Structural analysis of virulence factors and essential proteins of *Clostridium difficile*

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Nina Schwemmlein
aus Lichtenfels

Structural analysis of virulence factors and essential proteins of *Clostridium difficile*

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Nina Schwemmlein
aus Lichtenfels

1. Referent: Prof. Dr. Wulf Blankenfeldt
2. Referent: Prof. Dr. Michael Steinert
Eingereicht am: 23.05.2018
Mündliche Prüfung (Disputation) am: 20.09.2018

Druckjahr 2018

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Schwemmlein N*, Pippel J*, Gazdag E-M and Blankenfeldt W (2018) Crystal Structures of R-Type Bacteriocin Sheath and Tube Proteins CD1363 and CD1364 From *Clostridium difficile* in the Pre-assembled State. Front. Microbiol. 9:1750. (Figures that were adapted from this publication are indicated in the respective figure legend)
**Diese Autoren haben in gleichem Maße zu der Veröffentlichung beigetragen.*

Tagungsbeiträge

Schwemmlein, N., Gazdag, E. M., Gerhard, R. & Blankenfeldt, W.: Structural analysis of toxins from *Clostridium difficile*. (Poster). 5th Annual Retreat - Helmholtz Centre for Infection Research Graduate School, Goslar-Hahnenklee, Germany (May 2014).

Schwemmlein, N. & Blankenfeldt, W.: Expression and purification of toxins from *Clostridium difficile*. (Poster). 7th International PhD symposium of the Helmholtz International Graduate School for Infection Research, Braunschweig, Germany (December 2014).

Schwemmlein, N. & Blankenfeldt, W.: Expression, purification and structural analysis of Toxin B from *Clostridium difficile*. (Poster). 1st CSSB International Symposium, Hamburg, Germany (April 2015).

Popp, M., Sadana, P. & Schwemmlein, N.: (Towards) Elucidating molecular mechanisms of pathogens using X-rays. (Talk). 6th Annual Retreat - Helmholtz Centre for Infection Research Graduate School, Braunschweig, Germany (June 2015).

Schwemmlein, N. & Blankenfeldt, W.: Structural analysis of toxins from *Clostridium difficile*. (Poster). 8th International PhD symposium of the Helmholtz International Graduate School for Infection Research, Braunschweig, Germany (December 2015).

Karste, K. & Schwemmlein, N.: Recombinant protein expression for structural biology. (Talk). 7th Annual Retreat - Helmholtz Centre for Infection Research Graduate School, Goslar-Hahnenklee, Germany (June 2016).

Schwemmlein, N. & Blankenfeldt, W.: On the hunt – Structural analysis of essential proteins from *Clostridium difficile*. (Poster). 9th International PhD symposium of the Helmholtz International Graduate School for Infection Research, Braunschweig, Germany (December 2016).

Schwemmlein, N.: Structural analysis of essential proteins from *Clostridium difficile*. (Talk). 8th Annual Retreat - Helmholtz Centre for Infection Research Graduate School, Goslar-Hahnenklee, Germany (May 2017).

Schwemmlein, N.: Insights into diffocins – crystal structure of a phage-tail-like protein from *Clostridium difficile*. (Talk). 20th Heart-of-Europe Bio-Crystallography meeting, Wojanów Palace, Poland (Sepetmber 2017).

Index

Abbreviations

Å	Angström (0.1 nm)
APD	autoprotease domain
β-Me	β-Mercaptoethanol
BLAST	Basic Local Alignment Search Tool
bp	base pairs
CD	circular dichroism
CDI	*Clostridium difficile* infection
cPCR	colony polymerase chain reaction
CROP	combined repetitive oligopeptides
CV	column volume
Da	Dalton
ddH_2O	double destilled water
DESY	Deutsches Elektronen-Synchrotron
DNA	deoxyribonucleic acid
dNTPs	deoxynucleosid triphosphate
DPF	Dortmund Protein Facility at Max Planck Institute for Molecular Physiology, Dortmund, Germany
DTT	dithiothreitol
ε_λ	extinction coefficient at wavelength λ
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EM	electron microscopy
ESRF	European Synchrotron Radiation Facility
et al.	*et alii*
fwd	forward primer/oligonucleotide
g	unit of the gravitational acceleration
GTD	glucosyltransferase domain
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
HZI	Helmholtz Centre for Infection Research, Braunschweig, Germany
IEX	ion exchange chromatography
IPTG	isopropyl-β-D-1-thiogalactopyranoside
LB	lysogeny broth
MBP	maltose binding protein
MHH	Hannover Medical School
MR	molecular replacement
MWCO	molecular weight cut-off
Ni(II)-IMAC	Ni(II)-immobilized metal affinity chromatography

OD_{600}	optical density at 600 nm
ORF	open reading frame
p.a.	*pro analysi*
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pH	*potential hydrogenii*
Phyre²	Protein Homology/analogY Recognition Engine V 2.0
PMC	Pseudomembranous colitis
PSI	Paul Scherrer Institute
rev	reverse primer/oligonucleotide
revNi(II)-IMAC	reverse Ni(II)-immobilized metal affinity chromatography
r.m.s.d.	root mean square deviation
rpm	rounds per minute
SAD	single anomalous diffraction
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
SeMet	selenomethionine
SER	surface entropy reduction
SLIC	sequence-and-ligation-independent cloning
SLS	Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland
TAE	Tris acetate EDTA buffer
TB	terrific broth
TDSA	thermal denaturation shift assay
TEV	tobacco etch virus
T_m	melting temperature
Tris	tris (hydroxymethyl) amino methane

Amino acids are abbreviated with one letter and three letter codes according to the recommendations of the *International Union of Pure and Applied Chemistry* (IUPAC) and of the *International Union of Biochemistry and Molecular Biology* (IUB).

1. Summary

Clostridium difficile is a Gram-positive, anaerobic, endospore-forming bacterium that produces several virulence factors, most prominently the secreted protein toxins Toxin A (TcdA) and Toxin B (TcdB). *Clostridium difficile* infections (CDI) are often hospital acquired and antibiotic-associated. Treatment of CDI currently involves taking broad-spectrum antibiotics, e.g. vancomycin. Due to the extremely high relapse rate of CDI after antibiotic treatment, the emergence of new highly virulent *C. difficile* strains and the threatening antibiotic resistance, the need for new therapeutic treatment methods for CDI is more urgent than ever before. To develop new therapeutics, a detailed knowledge of the molecular processes inside the pathogen as well as a comprehensive structural and functional knowledge of its virulence factors and proteins involved in infection is essential. Aim of this thesis was therefore the structural characterization of the main virulence factor TcdB and of proteins that are involved in basic cellular processes, i.e. growth and sporulation, of *Clostridium difficile*.

The main virulence factor of *Clostridium difficile*, TcdB, belongs to the family of large clostridial toxins (LCTs) and consists of four functional domains, which mediate the mechanism of action. Via its receptor binding domain, the toxin binds to the host cell surface and is internalized via endocytosis. Acidification of the endosome causes conformational changes in the TcdB intermediate translocation domain, whereupon hydrophobic parts of this domain insert into the endosomal membrane and form a pore. Through this pore, the N-terminal glucosyltransferase domain is transferred into the host cell cytosol, where it inactivates host GTPases by glucosylation. Upon inactivation of these GTPases, several downstream signaling processes are disturbed, which can lead to inflammation response and cell death.
Other research groups have already published several structures of individual TcdB domains, e.g. the autoprotease and the glucosyltransferase domain. However, crystal structures of the translocation domain and of full-length TcdB were still missing. The aim of this project was a structural characterization of the translocation domain and full-length TcdB to get insights into the atomic structure and the translocation process of the virulence factor. Several constructs including full-length TcdB as well as truncated TcdB fragments were cloned for purification and crystallization experiments. However, protein crystallization of these constructs was not successful. In parallel, SAXS (small angle X-ray scattering) experiments of some TcdB constructs were performed to reveal the overall shape of the TcdB translocation domain. The calculated SAXS envelope shows high similarity to the TcdA translocation domain. This reveals that, besides their high sequence similarity, TcdA and TcdB share also a high structural similarity.

A second project of this thesis involved the functional and structural characterization of *Clostridium difficile* proteins that are essential for growth or sporulation of the pathogen *in vitro*

(Dembek *et al.*, 2015). Five functionally and structurally uncharacterized proteins (CD1067, CD1219, CD1823, CD2589 and CD2752) were selected with the objective of finding novel protein folds and potential drug targets for CDI treatment. Several constructs of all five proteins were cloned, expressed in *E. coli* and purified. Pure protein samples were used for crystallization experiments.

SAXS experiments revealed that CD1219 is present as a dimer in solution. The crystal structure of dimeric CD1219 could be determined by SeMet-SAD (single wavelength anomalous diffraction) to 1.8 Å resolution and reveals that the monomer of CD1219 consists of two domains. The large C-terminal domain shows structural similarities to P-loop GTPases and includes several secondary structure motifs and conserved residues that are essential for GTP binding. Its small N-terminal domain shares an OB-fold, which is characteristic for oligonucleotide-binding proteins, such as IF1s and cold-shock proteins. Nevertheless, further biochemical experiments are necessary to confirm DNA-/RNA- binding to the small N-terminal domain and GTP-binding to the large C-terminal domain of CD1219.

The crystal structure of CD1823 was solved by molecular replacement to 1.9 Å using the crystal structure of the H_2O_2 stress response protein YaaA from *E. coli* as a model. CD1823 shows a compact protein structure without characteristic conserved secondary structure motifs. Comparison with homologous proteins and analysis of the operon architecture of CD1823 reveals that the protein could be involved in peroxide stress response. SAXS analysis showed that it is monomeric in solution, but the structural analysis of CD1823 revealed no obvious hypotheses concerning its molecular function. Hence, further experiments are necessary to reveal the function of CD1823.

The crystal structures of the other three selected proteins (CD1067, CD2589, CD2752) could not be determined in this thesis due to challenges during protein purification, deficient crystal growth or insufficient diffraction quality of the protein crystals. Various attempts such as buffer optimization by TDSA (thermal denaturation shift assay), crystal optimization using grid screens or SER (surface entropy reduction) yielded only slight improvements.

Another part of this thesis dealt with the structural characterization of the diffocin proteins of *Clostridium difficile*. The diffocin particle is a contractile, bacteriophage-tail like bacteriocin. It consists of an outer contractile sheath and an inner non-contractile tube. Upon SOS-response, the diffocin particles are released by bacterial lysis, whereupon they can attack other *C. difficile* strains in the environment. After binding to the host cell, the diffocin sheath is contracted and the tube is pushed through the host cell membrane, which creates a small pore. This pore formation results in disturbance of the membrane potential, upon which the bacterial host cell is killed.

Both diffocin sheath and tube are built of multiple copies of one single protein, the sheath protein CD1363 and the tube protein CD1364. In this thesis, the crystal structures of both

proteins were determined to 1.9 and 1.5 Å resolution, respectively. Both proteins show high structural similarity to other bacteriophage-like proteins, e.g. the pyocin proteins of *Pseudomonas aeruginosa*. Based on homology of these related proteins, a model of the diffocin particle was created in this thesis.

2. Introduction

2.1. The pathogenic bacterium *Clostridium difficile*

Clostridium difficile is an anaerobic, Gram-positive, spore-forming bacterium that is widespread in nature and was first described in 1935 as an "actively motile, heavy-bodied rod with elongated subterminal or nearly terminal spores" (Hall & O'Toole, 1935). It can be found in the soil but is also present in the intestine of animals and humans. There, it can cause life-threatening infections, making it a huge hazard for the population. In 2011, 29.000 deaths associated with *Clostridium difficile* infections were registered in the USA (Center for Disease Control, U.S.). It is assumed that more than 65.000 CDI cases exist in Germany per year (Lübbert *et al.*, 2016). In the last years new epidemic *C. difficile* strains were detected in North America and Europe, which show increased toxicity and mortality (Kuijper *et al.*, 2006).

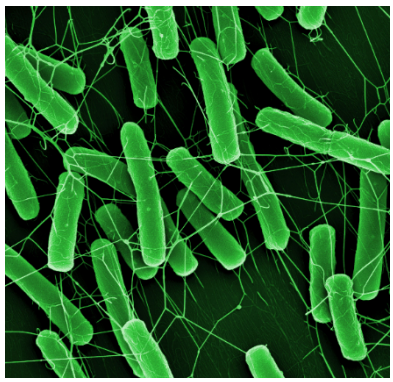

Figure 1: Electron microscopic image of *Clostridium difficile* (M. Rohde/HZI, M. Jahn/TU Braunschweig)

Most *Clostridium difficile* infections (CDI) are hospital-acquired or antibiotic-associated. Due to its spore-forming ability, the pathogen can survive on hospital surfaces for a long time, as *C. difficile* spores are extremely resistant to heat- or alcohol-based standard disinfectant treatments (Paredes-Sabja *et al.*, 2014). Furthermore, the spores and the bacterium can be transferred via the oral-fecal route between clinical staff and patients, which is a problem especially for elderly people (Martin *et al.*, 2016). As people > 65 years are more often in contact with hospital areas and the microbiome and physical condition changes upon age, they are more often infected with *C. difficile* than younger people (Jump, 2013).

Apart from long-term hospital stays, antibiotic treatments are an additional risk factor for CDI. Usually, the pathogen is outcompeted by other bacteria in the human gut. After disturbance of the microbiota, especially after treatment with broad-spectrum antibiotics (e.g. clindamycin), the majority of the human microflora is wiped out. Ingested spores can germinate and due to its resistance to antibiotics, *C. difficile* can colonize the human colon. The bacterium multiplies and releases its main virulence factors, the protein toxins, which cause major tissue damage and are the origin of the disease symptoms (Wilson, 1993).

Symptoms of *C. difficile* associated infections can range from mild diarrhea and fever to even life-threatening infections with a high death rate, such as toxic megacolon or pseudomembranous colitis (PMC) (Kuijper *et al.*, 2006). PMC is characterized by plaques (so-called pseudomembranes) in the mucosa of the colon (Bartlett *et al.*, 1978).

Due to the emergence of new highly virulent strains, the usually high antibiotic resistance of *C. difficile* and the high relapse rate of *C. difficile*-associated infections, there is an urgent need for new treatment methods (Smits *et al*., 2016).

2.2. Virulence factors of *Clostridium difficile*

The main virulence factors of *C. difficile* are the two secreted protein toxins, TcdA and TcdB. They belong to the family of large clostridial toxins (LCT) and are the causative agents of CDI (Awad *et al*., 2014). More details about the protein toxins of *Clostridium difficile* can be found in chapter 6.1.

Besides the homologous protein toxins, the pathogen possesses several further virulence factors. As it colonizes the gut of animals and humans, many different virulence factors are involved in adhesion to host cells and colonization processes. Among these are flagella (Delmee *et al*., 1990), which facilitate the accelerated and directed movement of the bacterium to the host cell and to the site of infection, as well as S-layer proteins (Karjalainen *et al*., 2001) and adhesins (Waligora *et al*., 2001), which enable adhesion to host-cell surfaces and colonization of the host.

2.3. Sporulation of *Clostridium difficile*

Anaerobic bacteria such as *C. difficile* usually cannot survive in an aerobic environment outside the host because they cannot tolerate the oxygenic environment. Also, in the host stomach the bacterium has to cope with a hostile environment due to the low pH. A route to endure these extreme conditions is to form resistant spores, which enable the bacterium to survive in these environments for a long time (Paredes-Sabja, *et al*., 2014). It could be shown that *C. difficile* strains that are unable to form spores are also not able to survive in an environment outside the host (Deakin *et al*., 2012). Furthermore, spores are resistant to antibiotics and attacks from the immune system (Ali *et al*., 2011). Thus, the spores of *Clostridium difficile* are considered to be the infectious form of the pathogen.

It is assumed that different environmental factors trigger sporulation in *Clostridium difficile*, such as starvation or quorum sensing (Higgins & Dworkin, 2012). In most *Clostridium* species, the regulation of sporulation involves phosphorylation of Spo0A, the master regulator of sporulation, by histidine kinases (Steiner *et al*., 2011). *C. difficile* spores are usually incorporated into the host via the oral-fecal route. They germinate in the host gut in presence of primary bile acids (e.g. taurocholate, cholate). This process is further supported by L-glycine (Sorg *et al*., 2008 and Wheeldon *et al*., 2011), which is recognized by the receptor CspC (Francis *et al*., 2013). The spore coat is then degraded due to proteolysis by serine proteases (Adams *et al*., 2013). This leads to the release of calcium dipicolinic acid, resulting in complete core hydration, outgrowth of vegetative cells and colonization of the host intestine (Paredes-

Sabja *et al*., 2014; Bhattacharjee *et al*., 2015). The vegetative cells can then produce toxins, the prominent virulence factors of *Clostridium difficile*. However, the exact mechanism and regulation of spore formation, germination and interaction with the host is still unknown.

2.4. Treatment of *Clostridium difficile*-associated infections

The standard therapy of *C. difficile* infections usually involves the administration of broad-spectrum antibiotics (e.g. metronidazole, vancomycin), but the relapse rate of CDI after taking these antibiotics is around 20% (Eyre *et al*., 2012). The infections reoccur usually 1-3 weeks after the antibiotic administration was stopped, but relapse after a few months has also been observed (Choi *et al*., 2011). The particular reasons for the recurrence of CDI are not known (Fekety *et al*., 1997), but risk factors for CDI recurrence are e.g. high age of patients (McFarland *et al*., 1999), use of fluorochinolones (Cadena *et al*., 2010), proton-pump inhibitors and a low level of serum albumin (Kim *et al*., 2010).

However, some new additional therapeutic treatment methods currently gain more and more popularity. One alternative to broad-spectrum antibiotics is the more specific antibiotic fidaxomicin (Vaishnavi, 2015). Furthermore, neutralizing human monoclonal antibodies are already in phase III clinical studies. Bezlotoxumab, a monoclonal TcdB-neutralizing antibody, has already been approved by the FDA (Food and Drug Administration) to support antibiotic treatment of CDI, in order to prevent recurrence of infection (Wilcox *et al*., 2017). By binding to the CROPs region of TcdB, Bezlotoxumab prevents toxin binding to host cell surfaces and neutralizes toxicity *in vitro* (Orth *et al*., 2014). Another promising therapy option is fecal microbiota transplantation. This technique involves the infusion of the stool of a healthy person in the gastrointestinal tract of a patient suffering from CDI. The disturbed microflora of the patient is regenerated by the microflora of a healthy person by defeating the pathogen. Unfortunately, long-term studies of this relatively new technique are currently not available and only few hospitals support this procedure (Lee *et al*., 2015; Kelly & Tebas, 2018).
Treatment of CDI might also involve small-molecule inhibitors (Slater *et al*., 2013). A few inhibitors are on their way and have already been FDA-approved. Among those is Ebselen, a compound that blocks the autoprotease domain of TcdB. A study shows reduced tissue damage in a mouse infection model upon infection with *C. difficile* and treatment with Ebselen (Bender *et al*., 2015). The compound further leads to inactivation of NADPH oxidase 1 activity (Smith *et al*., 2012), which reduces the toxin-induced production of reactive oxygen species. Another promising compound is the antioxidant *N*-acetylcysteine, which prevents TcdB-induced tissue damage in a colonic explant model (Farrow *et al*., 2013).

3. Research Objectives

Clostridium difficile is a pathogen that can cause life-threatening infections. Therapy of these infections is currently extremely problematic due to the ease of spread of the bacterium, its antibiotic resistance and the recurrence of infections after antibiotic treatment. Therefore, the need for new therapeutic treatment methods is very urgent. To develop new drugs, it is essential to obtain a detailed understanding of the molecular processes inside the pathogen and the mechanisms and structures of its virulence factors. To enlighten the details of molecular processes of *C. difficile*, one has to analyze the structure and function of essential proteins that are involved in central metabolic pathways. The aim of this thesis is therefore to characterize virulence factors, essential proteins and the diffocin proteins of *Clostridium difficile* structurally. The crystal structures can then be used to derive initial ideas about the molecular function of the respective proteins and can be considered as a basis for the development of new therapeutic treatment methods for CDI.

1. **Structural analysis of TcdB**

   The secreted protein toxin TcdB is one of the main virulence factors of *Clostridium difficile*. The crystal structures of some domains of TcdB have already been published, but the crystal structure of full-length TcdB and of its intermediate translocation domain were unknown at the outset of this thesis. To get a detailed insight into the translocation process and pore formation of the toxin, it is essential to determine the structure of its intermediate translocation domain. The crystal structure of full-length TcdB and its translocation domain could pave the way for designing new drugs that target the translocation domain specifically in order to inhibit toxin translocation and therefore prevent infection of the host. Consequently, one aim of this thesis was to determine the crystal structure of the intermediate TcdB domain by crystallographic methods, either with the full-length toxin or with fragments containing the entire intermediate domain or parts of this domain.

2. **Structural and functional studies of proteins involved in growth or sporulation of *Clostridium difficile***

   Structural and functional analysis of uncharacterized proteins that are involved in central metabolic processes can aid the identification of new potential drug targets. A recently published genome-wide study using transposon mutants identified several proteins that are involved in growth or sporulation of *Clostridium difficile in vitro* (Dembek *et al*., 2015). Since several of these proteins are structurally uncharacterized, they do most probably not belong to a common, well-characterized protein family. A

detailed structural characterization of these proteins could hence lead to identification of new potential drug targets for CDI therapy. Some of these uncharacterized proteins have therefore been selected for this thesis with the aim to characterize the proteins by crystallographic methods.

3. Structure determination of diffocin proteins

Diffocin particles are high-molecular weight bacteriocins that specifically kill other *C. difficile* strains in the environment of the diffocin-producing bacterium. Based on their architecture, diffocin particles share similarities to bacteriophage tails. They are assembled by several structural components and consist of a tail, which is composed of a contractile sheath and an inner rigid tube. The diffocin target specificity is determined by the receptor binding proteins. By engineering diffocins, one can direct the particles to target individual strains. Therefore, they are treated as a potential novel therapy for CDI. A structural characterization of the diffocin proteins can ease the process of diffocin engineering for therapeutic purposes. As structural data for the diffocin proteins have not been available at the beginning of this thesis, they should be characterized by crystallographic methods in this thesis.

4. Principles of X-ray crystallography

Protein X-ray crystallography was used for structure determination of proteins involved in disease and infection processes of *Clostridium difficile* in this thesis. The structural analysis of a protein can also help to derive ideas about its potential function and to get a detailed insight into its molecular mechanism. X-ray crystallography is therefore an important tool for the structural as well as the functional characterization of proteins.

4.1. Crystallization of proteins

To determine the three-dimensional structure of a protein by X-ray crystallography, the protein has to be crystallized. Protein crystallization is achieved by transition of the protein sample from the liquid into the crystalline phase. This transition is described by a so-called phase diagram, which is exemplarily shown in Fig. 2 (Asherie, 2004). The phase diagram is a two-dimensional plot of the protein concentration against precipitant concentration. It can be divided in three parts: the stable, metastable and unstable zone. In the stable zone, the protein is present in soluble form in solution. No crystal growth will take place. By transition above the solubility curve (blue line in Fig. 2), the protein reaches the metastable zone. This is typically achieved by incubation of the protein with precipitant solution, as this decreases the protein solubility. In the metastable region, spontaneous crystal growth can occur due to phase separation and spontaneous formation of nuclei. In the nucleation zone, supersaturation of the protein sample is high enough for crystal growth and spontaneous nucleation can take place. In the unstable region, also called precipitation zone, supersaturation (e.g. due to high precipitant concentration) of the sample leads to protein precipitation (Rupp, 2015).

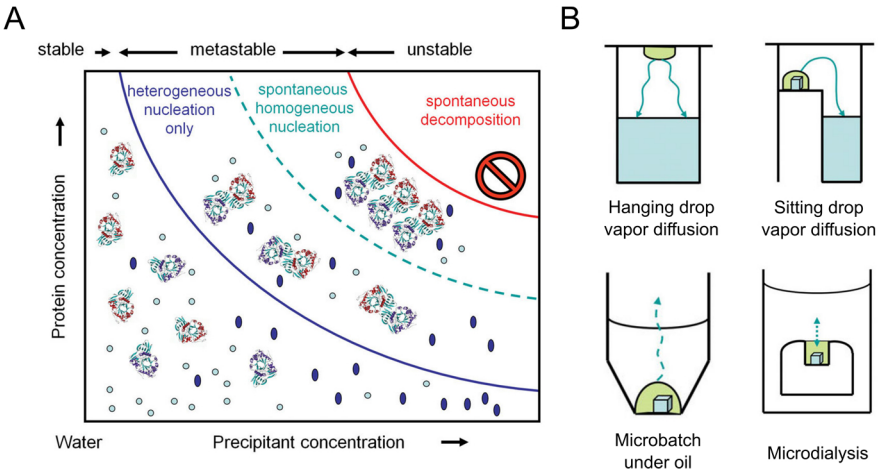

Figure 2: Schematic overview of protein crystallization. (A) A crystallographic phase diagram illustrates the state of the protein sample as a function of precipitant and protein concentration. The different zones (stable, metastable and unstable) are separated by a blue and red line, respectively. The blue line illustrates the solubility line and describes the transition of the protein from the soluble into the metastable zone. This transition is essential for nucleation and crystal formation. (B) Different methods of protein crystallization-set ups are depicted: hanging-drop and sitting-drop vapor diffusion, microbatch under oil and microdialysis. The figure is adapted from Rupp, 2009. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

Therefore, it is important to determine a suitable composition of precipitant solution and an appropriate crystallization setup for controlled crystal growth. Protein crystallization is influenced by several further parameters. An essential requirement for protein crystallization is a high sample purity and in most cases a high protein concentration (5 - 30 mg/ml). Temperature, pH, reservoir composition and concentration can also affect crystal growth. There are several different crystallization setups for protein crystallization. In this study, the sitting-drop vapor-diffusion method was used. Protein and reservoir (precipitant) solution are usually mixed in a 1:1 ratio and incubated in a chamber above the precipitant-containing reservoir solution as a “sitting drop”. As the protein drop contains less precipitant solution compared to the reservoir, water diffuses from the protein drop into the reservoir solution in a closed system to compensate this imbalance. Hence, the protein and precipitant concentration in the drop increase and transition of the protein from the soluble into the metastable phase is initiated (Rupp, 2009).

As it is dependent on many different parameters, protein crystallization is an unpredictable process. It can sometimes take weeks or months, but does not have to be successful at all. It is therefore important to monitor the crystallization setup continuously over time until crystal growth is observed.

4.2. Diffraction of X-rays and data collection

Once single protein crystals of suitable size (50 - 500 µm) are obtained, diffraction experiments can be performed. Prior to data collection, cryoprotection of the crystal is important to avoid ice formation, which can result in destruction of the crystal arrangement and can lead to loss of diffraction quality (Haas & Rossmann, 1970; Garman, 2003). The use of cryoprotectants, such as glycerol or 2,3-butanediol, is usually dependent on the reservoir composition and the amount of cryoprotectant solution has to be determined for each crystal individually. However, some crystallization conditions do not require the use of additional cryoprotectants, as latter are already included in the reservoir solution. The crystals can then be fished directly from the crystallization screen using nylon loops and need to be flash-cooled in liquid nitrogen immediately.

After cryoprotection and harvesting, the crystal is mounted in the X-ray beam on a goniometer in front of the X-ray source. A schematic experimental setup is illustrated in Fig. 3. It usually includes an X-ray source for generation of a focused X-ray beam of specific wavelength and a goniometer head for crystal mounting. A cryocooling system with a nitrogen gas stream at 100 K protects the crystal from radiation damage upon formation of radicals during data collection. The detector, which is required for collection of the diffraction pattern, and a beam stop, which prevents the detector from damage by non-diffracted X-rays, are located behind the goniometer (Rhodes, 2006).

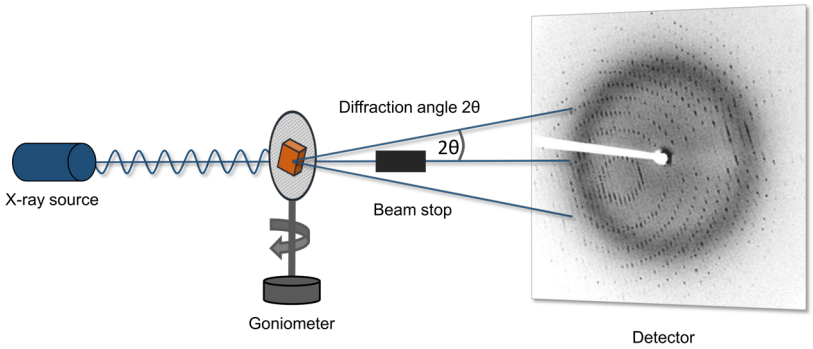

Figure 3: Experimental setup of an X-ray diffraction experiment. The X-ray source produces a focused X-ray beam of a certain wavelength that hits the crystal, which is mounted on a goniometer head. The diffracted X-rays are recorded on a detector. A beam stop is essential to prevent damage of the detector by non-diffracted intense X-rays.

X-rays of a certain wavelength are used for protein X-ray diffraction experiments. The resolution of a protein structure is determined by the minimal distance between two points that can be clearly separated from each other and is dependent on the wavelength of the used light. As the bond length between two atoms is usually in the range of a few Å, X-rays with a wavelength of ~ 0.1 - 100 Å are used for X-ray crystallography to resolve a protein structure at the molecular level (Rupp, 2009).

As X-rays are only weakly diffracted by a single protein molecule, the diffraction signal has to be amplified. This is achieved by the use of a protein crystal, which includes identical copies of the protein in a regular periodic arrangement. When the X-rays hit the protein crystal, they are scattered by the electrons of the protein atoms, which are located on the lattice planes of the crystal. It is therefore assumed that the incoming X-rays are reflected at these planes (Fig. 4). Diffraction of the X-rays can be described by Bragg's law (equation 1; Bragg, 1913).

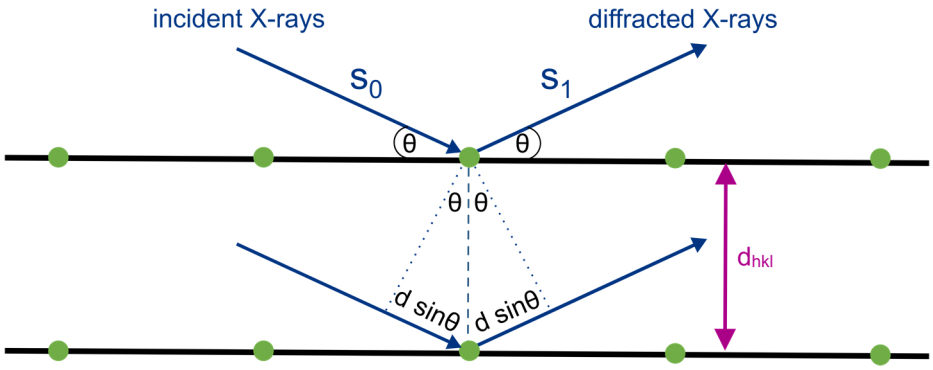

Figure 4: Schematic illustration of Bragg's law. Two incident X-rays are reflected at the crystal planes with the distance d_{hkl} by the angle θ. Constructive interference occurs, when d_{hkl} is an integer of the wavelength λ.

$$n \cdot \lambda = 2d \cdot \sin\theta \qquad (1)$$

In this equation, n is an integer, λ is the wavelength of the incoming electromagnetic wave, d is the distance between the planes in the crystal and θ is the diffraction angle of the X-ray.

When *d* is a positive integer of the wavelength *λ*, diffraction can be observed. This phenomenon is also called constructive interference. The scattered waves interfere constructively and the sum of their reflections is recorded on the detector. The waves that interfere destructively are cancelled out. This leads to a distinct diffraction pattern that consists of many single spots, the reflections. It contains information about the unit cell and spacegroup of the protein crystal. The intensity of the reflections gives information about the electron envelope of the protein.

The diffraction pattern that is visible on the detector can be explained by transcribing Bragg's law into the reciprocal space, which is inversely proportional to the real space. For this purpose, an Ewald sphere has to be calculated (Ewald, 1913). This is a three-dimensional sphere with the radius $r = 1/\lambda$ centered on the crystal and is exemplarily illustrated in Fig. 5. The origin of the reciprocal crystal lattice is located in the intersection of the Ewald sphere with the direct X-ray beam. Bragg's law is true for each lattice point that is located on the surface of the Ewald sphere. This indicates constructive interference and the respective point will be visible as a reflection on the detector. Due to rotation of the crystal during data collection, various lattice points will lie on the Ewald sphere and produce reflections on the detector. The diffraction pattern is therefore a direct projection of the three-dimensional reciprocal crystal lattice. Each diffraction point is described by its intensity, the Miller indices *h*, *k*, *l* of the reciprocal crystal lattice and is dependent on the diffraction angle *θ* (Foadi & Evans, 2008).

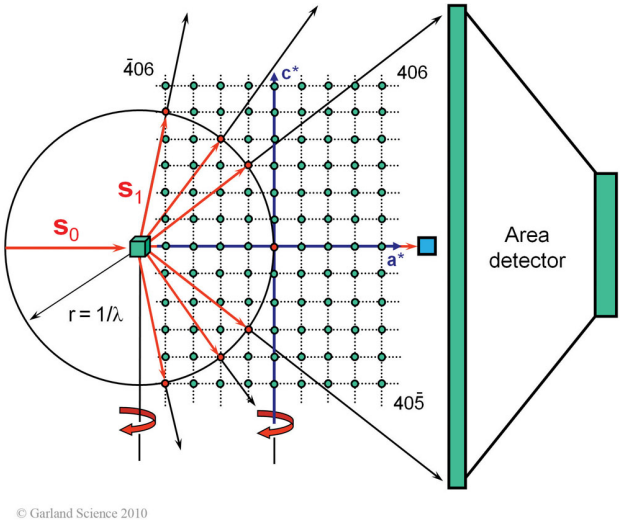

Figure 5: Schematic representation of the Ewald sphere. The construction of an Ewald sphere (sphere with radius $1/\lambda$ around the crystal) explains the diffraction pattern on the detector. In this figure, the protein crystal is depicted as a green cube, incident X-rays are diffracted by the crystal (red arrows). The reciprocal crystal lattice is illustrated by black dashed lines, the lattice points as green dots. Whenever a lattice point lies on the Ewald sphere (red dot), Bragg's law is satisfied and this point is visible as a reflection on the detector due to constructive interference. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

4.3. Data processing and structure determination

The diffraction pattern of the protein crystal is recorded on the detector during the diffraction experiment. The positions and intensities of the reflections have to be determined to calculate the desired electron density $\rho(r)$, which is essential for structure determination.

The processing of the dataset starts with indexing of the reflections. During this step, the unit cell parameters and the space group of the protein crystal are determined. Therefore, one needs to extract the reflection positions, as well as experimental data such as detector distance, wavelength and beam position from the dataset. Crystal orientation and possible spacegroups are proposed and the Miller indices for each reflection are assigned to describe the correct position of each reflection (Powell, 2017).

The asymmetric unit (AU) is the smallest element that describes a crystal. By rotation and translation of the AU, the unit cell is created, which is characterized by one of the 14 Bravais lattices (Fig. 6). The lattice symmetry of the crystal is in turn defined by one of the 65 possible spacegroups. They describe the respective symmetry operators that are applied to construct the unit cell (Rupp, 2009).

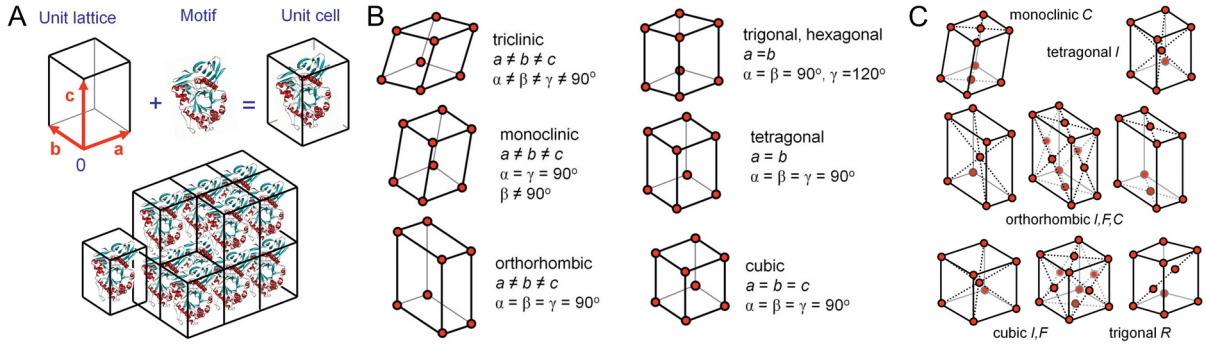

Figure 6: Organization of protein molecules in a crystal. (A) The unit cell is described by the unit lattice, which is filled with a protein motif. A crystal contains several copies of this unit cell that are arranged in the crystal lattice. (B) Overview of the six primitive three-dimensional crystal lattice types that are described by the vectors *a*, *b*, *c* and the angles *α*, *β*, *γ*. (C) Overview of the eight centered three-dimensional crystal lattice types, which result in 14 possible Bravais-lattices in total together with the six primitive types. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

During integration, the second step of data processing, the intensity of a specific region is measured in contrast to the background. Due to the crystal symmetry, one reflection will be measured several times concerning the different crystal orientations during rotation of the crystal in the beam. The next step of data processing is the reduction step. Thereby, the dataset is reduced to only one single intensity value per reflection. The following scaling step yields an averaged intensity value for each reflection, considering crystal morphology and beam intensity (Powell, 2017).

The electron density can then be calculated using the following equation:

$$\rho(xzy) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{\mathrm{hkl}} \cdot e^{i\alpha_{\mathrm{hkl}}} \cdot e^{-2\pi i(\mathrm{h}x+\mathrm{k}y+\mathrm{l}z)} \quad (2)$$

In equation 2, V is the unit cell volume, F_{hkl} is the structure factor amplitude for each reflection and α_{hkl} its associated phase. The electron density can hence only be calculated when amplitude and phase of each reflection are known. The amplitudes can be derived directly from the experimentally measured intensities I_{hkl}, as the intensities are proportional to the structure factor amplitudes ($I = |F_{hkl}|^2$). However, the phase information is lost during the diffraction experiment. This is the so-called phase problem of X-ray crystallography (Rupp, 2009).

Several methods can be used to solve the phase problem. Phases can for example be determined by Molecular Replacement (MR). For this method, one needs a protein model that shares a high sequence similarity with the uncharacterized protein. It is assumed that proteins that share at least 30% sequence identity adopt a similar fold. The structure of a similar protein can hence be used as a search model to solve the structure of a protein with unknown structure by MR. The phases of the search model will thereby give an estimation of the phases of the unknown protein. For MR, a dataset of the native protein crystal is collected and the unit cell is determined by processing the dataset analogous to the above description of data processing. The structure of the known protein is then placed in the unit cell of the unknown protein by rotational and translational search. Once in the correct position and orientation, phases of the known structure and intensities of the unknown structure are combined to calculate the electron density for the unknown protein structure (Evans & McCoy, 2007).

When no suitable MR model is available for structure determination by MR, phases can also be determined experimentally. For experimental phasing, heavy atoms are introduced into the protein by expression of SeMet-labeled protein or by soaking heavy atoms, e.g. mercury or gold, into the protein crystal. In this study, Se-SAD was used for experimental phasing. SAD is short for "single-wavelength anomalous diffraction". For this method, heavy atoms that absorb X-rays at a specific wavelength (absorption edge) are used, as the intensity and phase of the diffracted X-rays are influenced in the presence of an anomalous scatterer. In detail, the atomic scattering factor f is impacted by the two additional components f' and f'' in the presence of heavy atoms.

$$f = f_n + \Delta f' + i\Delta f'' \quad (3)$$

In equation 3, f_n is the normal scattering factor of the native protein, which is dependent on the diffraction angle, but independent of the wavelength. $\Delta f'$ is the dispersive component, which describes the reduction of the measured intensities at different wavelengths. $\Delta f''$ is the imaginary component, which shows a 90° phase shift compared to the normal and dispersive

component. In contrast to the normal component, the two factors *f'* and *f''* are dependent on the wavelength that is used for the diffraction experiment. The influence of the anomalous scattering factors on the native scattering factor are illustrated in Fig. 7B. This anomalous difference is used to locate the heavy atoms in the unit cell by Patterson functions.

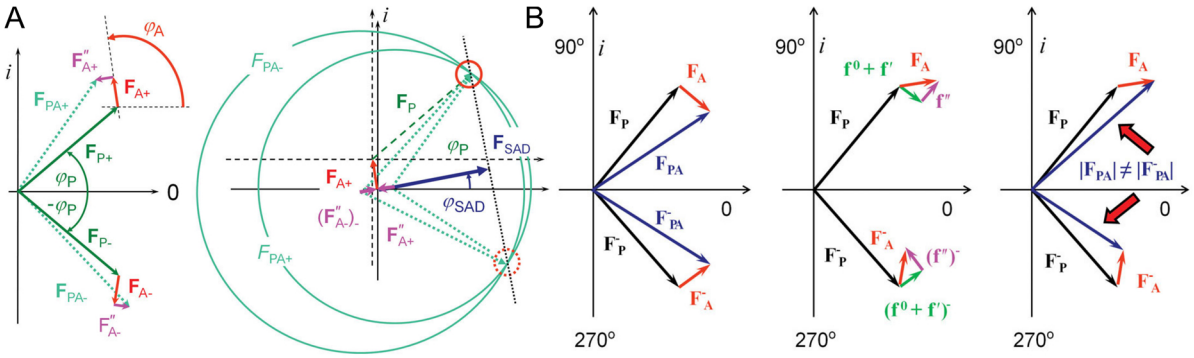

Figure 7: Impact of an anomalous scatterer on the structure factor amplitude and its use for the determination of the phase problem. (A) F_{PA} is the structure factor of the anomalous derivative of the protein, F_A that of the anomalous scatterer and F_P that of the native protein. The amplitude of F_{PA} is measured experimentally and the position of the anomalous scatterers can be determined by Patterson functions. Two possibilities for the phase angle of the protein are thus possible, indicated by red circles in the right panel. (B) Illustration of Friedel's law. When Friedel's law is true, two structure factors F_P and F_{-P} are real functions and have the same amplitude, but inverse phases (left panel). F_{PA} is therefore equal to its Friedel pair F_{-PA}. In the presence of an anomalous scatterer, Friedel's law is not true. The phases of the scattering factors *f'* and *f''* are changed, as the imaginary component *f''* of the scattering factor shows a phase shift of 90° (middle panel). F_{PA} is hence not equal to F_{-PA} anymore (right panel). Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

The structure factors contain information of all atoms that are present in the crystal. The structure factor of a protein crystal that contains heavy atoms (PA) is thus described by the sum of the structure factors of the protein atoms (P) and heavy atoms (A), $F_{PA} = F_P + F_A$. According to Friedel's law, the structure factor F_{hkl} and its Friedel mate F_{-h-k-l} have the same amplitude, but opposite phases. In the presence of an anomalous scatterer, Friedel's law is not true anymore, as the Friedel pairs have different amplitudes and different phases ($F_{hkl} \neq F_{-h-k-l}$). Fig. 7 illustrates the impact of an anomalous scatterer (heavy atom) on the structure factors (Rupp, 2009). In Fig. 7B, the structure factor for the native protein (F_P) is shown in black, that of the anomalous protein derivative (F_{PA}) in blue and the structure factor of the heavy atom (F_A) in red. For the native protein, Friedel's law is true and the Friedel pair F_{PA} and F_{-PA} show the same amplitude and inverse phases (mirrored at the real axis), as illustrated in the left panel of Fig. 7B. For the anomalous protein derivative (right panel of Fig. 7B), the scattering is influenced by the heavy atoms, which were introduced into the native protein, resulting in a phase shift and breakdown of Friedel's law. The structure factor amplitudes for the Friedel pair are hence not equal anymore (indicated by different lengths of the vectors F_{PA} and F_{-PA}). By combining the information of the measured amplitudes and the location of the anomalous scatterer in the protein, two phase solutions are possible (highlighted by red circles in Fig. 7A), but only one of the two possible phase angles is correct. Iterative cycles of density modification techniques such as solvent flattening can then be used to

determine the correct phase angle and to solve the phase problem. The correct phase solution is also indicated by decreasing R-values during refinement.

4.4. Model building, refinement and validation of the solved structure

Once an electron density map is calculated, the amino acid chain can be built into the electron density map by automated or manual model building. An initial model is built concerning different parameters, such as secondary structure constraints, length and angle of chemical bonds and distances between neighboring backbones and side chains. Depending on the resolution of the structure, specific details of the amino acid side chains can be visualized in the electron density map. The better the resolution, the more details can be built, e.g. different amino acid rotamers. Water molecules or ligands can also be visualized in the electron density map with increasing resolution (Rhodes, 2006).

After each model building step, a refinement of the model against the experimental data is performed using an algorithm. The theoretical structure factors of the model (F_{calc}) are calculated and compared to the experimentally measured structure factors (F_{obs}). During model building, the difference between these structure factors should be minimized to yield a good fit of model data and experimental data. The quality of this fit is described by *R*-factors, which illustrate the correlation between model (F_{calc}) and experimental data (F_{obs}). Two *R*-factors are usually considered: R_{work} and R_{free} (Brünger, 1992).

$$R_{work} = \frac{\sum_{hkl} |F_{obs} - kF_{calc}|}{\sum_{hkl} F_{obs}} \quad (4)$$

R_{work} directly correlates the model with the experimental data, while R_{free} is calculated excluding 5% of the reflections to avoid overfitting of the data. Both *R*-factors should decrease during refinement to indicate that the previous step of model building was successful. When the process of model building is finished, both *R*-factors should not change anymore during refinement and should not differ by more than 5% (Brünger, 1992).

The quality of the final structure is usually validated by different parameters and plots. The Ramachandran plot illustrates the statistical distribution of peptide torsion angles φ and ψ that cluster in specific regions for the respective secondary structure motifs that are present in the protein structure. Amino acids that show unusual peptide bond torsion angles are clustered in so-called “unallowed” regions of this plot (Ramakrishnan & Ramachandran, 1965).

Bond lengths and angles are compared to empirical values and are described by the r.m.s.d. from the empirical values. They should not exceed a certain threshold. Another parameter that can be used for quality analysis is the MolProbity score. It is an overall score for the evaluation of the structure quality, as it combines several criteria for quality analysis (e.g. rotamer outliers, torsion-angle analysis etc.) in one single score (Davis *et al.*, 2004).

5. Materials and Methods

5.1. Chemicals, solutions and materials

5.1.1. Chemicals

Unless stated differently, all chemicals were purchased at the highest purity available (p.a., *pro analysi*) from the companies AppliChem (Darmstadt, Germany), Fluka (Seelze, Germany), Merck (Darmstadt, Germany), Qiagen (Hilden, Germany), Roth (Karlsruhe, Germany), Sigma-Aldrich (Taufkirchen, Germany) and ThermoFisher Scientific Inc. (Waltham, Massachusetts, USA).

5.1.2. Buffers, media and stock solutions

Buffers and stock solutions were prepared in ddH_2O and filter sterilized before use; buffers for protein preparation were additionally degassed before use. Media were autoclaved. Antibiotics were used at concentrations of 100 µg/ml ampicillin, 34 µg/ml chloramphenicol and 5 µg/ml tetracycline.

Solutions for agarose gel electrophoresis:

50x TAE buffer	242 g Tris, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA pH 8.5 in 1 L ddH_2O
0.8% agarose solution	2 g agarose in 250 ml 1x TAE buffer
DNA standard	GeneRuler DNA Ladder Mix, Thermo Scientific
DNA sample buffer	20% Ficoll 400, 0.25% bromophenol blue, 0.25% xylenecyanol

Media:

LB agar	10 g/L NaCl, 10 g/L bacto tryptone, 5 g/L yeast extract, 15 g/L agar
LB medium	10 g/L NaCl, 10 g/L bacto tryptone, 5 g/L yeast extract
TB medium	12 g/L bacto tryptone, 24 g/L yeast extract, 0.4% (w/v) glycerol, 170 mM KH_2PO_4, 720 mM K_2HPO_4
SOC agar	20 g/L bacto tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, pH 7.5, 15 g/L agar
SOC medium	20 g/L bacto tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, pH 7.5
5x M9 salts	64 g/L Na_2HPO_4 x 7 H_2O, 15 g/L KH_2PO_4, 5 g/L NH_4Cl, 2.5 g/L NaCl in 1 L ddH_2O
M9 Minimal medium	780 ml autoclaved ddH_2O, 200 ml 5x M9 salts, 20 ml 20% α-D-glucose, 1 ml 2 M $MgSO_4$, 100 µl 1 M $CaCl_2$
Autoinduction medium	TB-medium, 0.01% lactose, 2 mM $MgSO_4$

Buffers for protein purification:

Strep-TcdB543-2366(C698A)-His_6

Ni(II)-IMAC Buffer A	20 mM Tris/HCl pH 8.0, 300 mM NaCl
Ni(II)-IMAC Buffer B	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 500 mM imidazole
Dialysis Buffer	20 mM Tris/HCl pH 8.0, 300 mM NaCl
Strep-Trap Buffer A	20 mM Tris/HCl pH 8.0, 300 mM NaCl
Strep-Trap Buffer B	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 5 mM Desthiobiotin
SEC buffer	20 mM Tris/HCl pH 8.0, 300 mM NaCl

TcdB800-1834

Ni(II)-IMAC Buffer A	20 mM Tris/HCl pH 8.0, 500 mM NaCl, 10 mM imidazole
Ni(II)-IMAC Buffer B	20 mM Tris/HCl pH 8.0, 500 mM NaCl, 500 mM imidazole
MBP affinity Buffer A	20 mM Tris/HCl pH 8.0, 500 mM NaCl
MBP affinity Buffer B	20 mM Tris/HCl pH 8.0, 500 mM NaCl, 10 mM maltose
Dialysis Buffer	20 mM Tris/HCl pH 8.0, 500 mM NaCl
IEX Buffer A	20 mM Tris/HCl pH 8.0, 20 mM NaCl
IEX Buffer B	20 mM Tris/HCl pH 8.0, 1 M NaCl
SEC buffer	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 10% glycerol

native CD1219

Ni(II)-IMAC Buffer A	20 mM Tris/HCl pH 8.0, 300 mM NaCl
Ni(II)-IMAC Buffer B	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 500 mM imidazole
Dialysis Buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl
IEX Buffer A	20 mM Tris/HCl pH 8.0, 20 mM NaCl
IEX Buffer B	20 mM Tris/HCl pH 8.0, 1 M NaCl
SEC buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl

SeMet-CD1219
SeMet-CD2589

Ni(II)-IMAC Buffer A	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 1 mM DTT
Ni(II)-IMAC Buffer B	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 1 mM DTT, 500 mM imidazole
Dialysis Buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl, 2 mM DTT
IEX Buffer A	20 mM Tris/HCl pH 8.0, 10 mM NaCl, 2 mM DTT
IEX Buffer B	20 mM Tris/HCl pH 8.0, 1 M NaCl, 2 mM DTT
SEC buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl, 2 mM DTT

native CD1823

Ni(II)-IMAC Buffer A	20 mM Tris/HCl pH 8.0, 300 mM NaCl
Ni(II)-IMAC Buffer B	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 500 mM imidazole
Dialysis Buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl
SEC buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl

native CD1363
native CD1364

MBP affinity chromatography Buffer A	20 mM HEPES pH 7.5, 300 mM NaCl
MBP affinity chromatography Buffer B	20 mM HEPES pH 7.5, 300 mM NaCl, 10 mM maltose
Dialysis Buffer	20 mM HEPES pH 7.5, 150 mM NaCl
Ni(II)-IMAC Buffer A	20 mM HEPES pH 7.5, 150 mM NaCl
Ni(II)-IMAC Buffer B	20 mM HEPES pH 7.5, 300 mM NaCl, 500 mM imidazole
SEC buffer	20 mM HEPES pH 7.5, 150 mM NaCl

native CD2589-SER1

Ni(II)-IMAC Buffer A	20 mM Tris/HCl pH 8.0, 500 mM NaCl
Ni(II)-IMAC Buffer B	20 mM Tris/HCl pH 8.0, 500 mM NaCl, 500 mM imidazole
Dialysis Buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl
IEX Buffer A	20 mM Tris/HCl pH 8.0, 15 mM NaCl
IEX Buffer B	20 mM Tris/HCl pH 8.0, 1 M NaCl
SEC buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl

native CD2752

Ni(II)-IMAC Buffer A	20 mM Tris/HCl pH 8.0, 500 mM NaCl, 2 mM DTT
Ni(II)-IMAC Buffer B	20 mM Tris/HCl pH 8.0, 500 mM NaCl, 500 mM imidazole, 2 mM DTT
Dialysis Buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl, 2 mM DTT
SEC buffer	20 mM Tris/HCl pH 8.0, 150 mM NaCl, 2 mM DTT

Solutions for SDS-PAGE:

4x stacking gel buffer	0.5 M Tris/HCl pH 6.8, 0.4% (w/v) SDS
Stacking gel (5%)	1x stacking gel buffer, 5% (w/v) acrylamide (from 30% (w/v) stock solution: 37.5:1), 0.1% (w/v) APS, 0.001% (v/v) TEMED
4x resolving gel buffer	1.5 M Tris/HCl pH 8.8, 0.4% (w/v) SDS
Resolving gel (15%)	1x resolving gel buffer, 15% (w/v) acrylamide (from 30% (w/v) stock solution: 37.5:1), 200 µL 10% APS, 0.001% (v/v) TEMED
8x SDS-PAGE sample buffer	2.2 ml 100 mM Tris/HCl pH 6.8, 10% (w/v) SDS, 4 ml (v/v) 100% glycerol, 800 µl β-Me, 0.05% bromophenol blue
Running buffer	25 mM Tris/HCl, 192 mM glycine, 0.1% (w/v) SDS
Protein standard	Protein Marker I, 14.4-116 kDa, VWR PageRuler™ Plus Prestained Protein Ladder, 10-250 kDa, ThermoFisher Scientific

5.1.3. Oligonucleotides and DNA templates

All oligonucleotides used in this study were designed using the program OligoCalc (Kibbe, 2007) and were purchased from MWG Eurofins Genomics (Munich, Germany). Sequences of the open reading frames (ORFs) were picked from the UniProt (The UniProt Consortium, 2017) database. A detailed list of oligonucleotides used in this study can be found in table 31 – table 37 in the supplement.

Different DNA templates were used for gene amplification and cloning in this study. Some templates were purchased as synthetic DNA constructs codon-optimized for expression in *E. coli*. Other constructs were cloned based on genomic *Clostridium difficile* DNA, a kind gift from Prof. Dr. Ralf Gerhard (MHH, Hannover, Germany). Some TcdB constructs were created based on the plasmid pHIS1522-TcdB543-2366 (kind gift from Prof. Dr. Ralf Gerhard, MHH, Hannover) as DNA template. Other TcdB fragments were cloned based on TcdB-DNA Codon-optimized for expression in *E. coli*, which was ordered at GenScript (Piscataway, USA). Synthetic DNA templates ordered at Life Technologies GmbH (Darmstadt, Germany) were used for amplification of the uncharacterized proteins essential for growth or sporulation of *Clostridium difficile*. All diffocin proteins were amplified from genomic *Clostridium difficile* DNA.

5.1.4. Plasmids and constructs

All plasmids and constructs that were used or generated in this study are listed in table 38 in the supplement.

5.1.5. Microorganisms

Different *E. coli* strains were used for amplification of constructs and heterologous protein expression in this thesis, respectively. All strains are listed in table 1.

Table 1: Overview of all microorganisms used in this study.

Bacterial strain	Genotype/Origin	Antibiotic resistance	Used for
Bacillus megaterium	Prof. Dr. Ralf Gerhard, MHH, Hannover	Tetracycline	Protein expression (TcdB)
E. coli Omnimax	F´ {proAB lacI^q lacZΔM15 Tn10(Tet^R) Δ(ccdAB)} mcrA Δ(mrr hsdRMS-mcrBC) Φ80(lacZ)ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD	Tetracycline	Cloning / amplification of plasmids
E. coli TOP10	F$^-$ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galKrpsL (StrR) endA1 nupG	Streptomycin	Cloning / amplification of plasmids
E. coli XL1 blue	F´ proAB lacIqZΔM15 Tn10 Tet^r/ recA1 endA1 gyrA96thi-1 hsdR17 supE44 relA1 lac	Tetracycline	Cloning / amplification of plasmids
E. coli Rosetta2(DE3)	F$^-$ ompT hsdS$_B$(r_{B^-} m_{B^-}) gal dcm (DE3) pRARE2 (Cam^R)	Chloramphenicol	Protein expression
E. coli BL21 CodonPlus(DE3)-RIL	F$^-$ ompT hsdS(r_{B^-} m_{B^-}) dcm$^+$ Tet^r gal λ(DE3) endA Hte [argU ileYleuW Cam^r]	Chloramphenicol	Protein expression

5.1.6. Enzymes

All restriction enzymes used for cloning in this study (BamHI, DpnI, HindIII, KpnI, NdeI, XhoI), as well as Antarctic Phosphatase, T4 DNA Polymerase, Recombinase A and T4 DNA Ligase were purchased from New England Biolabs (Ipswich, MA, U.S.A.). Phusion DNA Polymerase was ordered at ThermoFisher (Waltham, Massachusetts, USA), Taq Polymerase at VWR (Darmstadt, Germany) and KAPAHiFi Polymerase at Peqlab (Erlangen, Germany).

TEV protease was produced in the laboratory from plasmid pHR-MBP_TEV (DPF, MPI Dortmund) as His_6-MBP-TEV fusion protein and was purified via Ni(II)-immobilized metal ion affinity chromatography (Ni(II)-IMAC). PreScission protease was expressed from plasmid pET28(+) and was used after purification by Ni(II)-IMAC and size exclusion chromatography. Sumo protease was produced from plasmid pXX-Sumo and was purified by Ni(II)-IMAC and size exclusion chromatography.

5.1.7. Columns for protein purification

All columns that were used for protein purification in this study are listed in table 2.

Table 2: Overview of columns that were used for protein purification in this study.

HisTrap™ HP 5 mL column	GE Healthcare, Munich, Germany
HiTrap™ TALON® crude 5 mL column	GE Healthcare, Munich, Germany
HiTrap™ Q FF 5 mL column	GE Healthcare, Munich, Germany
StrepTrap™ HP 5 mL column	GE Healthcare, Munich, Germany
Strep 10 mL column (column material: Strep-Tactin® Superflow®, IBA GmbH, Göttingen, Germany)	prepared by Peer Lukat, HZI, Braunschweig, Germany
MBPTrap™ HP 5 mL column	GE Healthcare, Munich, Germany
HiLoad™75 16/60 Superdex™75 prepgrade	GE Healthcare, Munich, Germany
HiLoad™75 26/60 Superdex™75 prepgrade	GE Healthcare, Munich, Germany
HiLoad™200 16/60 Superdex™200 prepgrade	GE Healthcare, Munich, Germany
HiLoad™200 26/60 Superdex™200 prepgrade	GE Healthcare, Munich, Germany
Superose® 6 10/300 GL	GE Healthcare, Munich, Germany
Superdex™ 75 10/300 GL	GE Healthcare, Munich, Germany
Superdex™ 200 10/300 GL	GE Healthcare, Munich, Germany

5.1.8. Kits

All commercially available molecular biology kits and crystallographic kits that were used in this study are summarized in table 3.

Table 3: Overview of molecular biology kits and crystallography screens used in this study.

Molecular Biology Kits	
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
QIAquick PCR Purification Kit	Qiagen, Hilden, Germany
QIAprep Spin Miniprep Kit	Qiagen, Hilden, Germany
Crystallographic Kits / Screens	
Index™	Hampton Research, Aliso Viejo, USA
JCSG+, JCSG Core I, II, III and IV	Qiagen, Hilden, Germany
MIDAS™	Molecular Dimensions, Suffolk, U.K.
Morpheus®	Molecular Dimensions, Suffolk, U.K.
Protein Complex Suite	Qiagen, Hilden, Germany

5.2. Cloning

In this study, two different methods were used for cloning: traditional cloning using restriction enzymes as well as sequence-and-ligation-independent cloning (SLIC).

5.2.1. Traditional cloning

5.2.1.1. Polymerase chain reaction

Polymerase chain reactions (PCR) were conducted for amplification of target DNA. For this purpose, genomic *Clostridium difficile* DNA (CD630, Prof. Dr. Ralf Gerhard, MHH, Hannover, Germany), as well as synthetic DNA codon-optimized for expression in *E. coli* were used as DNA templates. In this study, *Phusion* DNA polymerase and *KAPAHifi* DNA polymerase were used for PCR experiments. *Taq* DNA polymerase was used for colony PCR (cPCR). A pipetting scheme for the respective PCR reactions and corresponding PCR programs are shown in table 4 and table 5. After addition of DNA polymerase to the PCR mixture, PCR programs were started immediately. 30 cycles of denaturation, annealing and elongation were used for standard PCR.

Table 4: Pipetting scheme for PCR reactions.

	Phusion Polymerase	*KAPAHifi* Polymerase	*Taq* Polymerase	final concentration
Template DNA*	x µl	x µl	1 µl	ca. 100 ng
Primer fwd	2.5 µl	0.5 µl	0.5 µl	2-5 µM
Primer rev	2.5 µl	0.5 µl	0.5 µl	2-5 µM
dNTPs	1 µl	0.75 µl	-	200-300 µM
Buffer	10 µl GC buffer	5 µl GC buffer	-	1x
ddH_2O	to 50 µl	to 25 µl	5.5 µl	
Polymerase	1 µl	0.5 µl	7.5 µl RedTaq Mastermix	

* For genomic DNA as template more DNA is required. For cPCR one bacterial colony was resuspended in 10 µl sterile ddH_2O and used as DNA template for cPCR.

Table 5: Overview of PCR programs.

		Phusion Polymerase		*KAPAHifi* Polymerase		*Taq* Polymerase	
		T (°C)	t (s)	T (°C)	t (s)	T (°C)	t (s)
	Initial denaturation	98	180	95	180	95	300
30x	Denaturation	98	10	98	20	95	30
	Annealing	T_m – 5 °C	30	T_m – 5 °C	15	T_m – 5 °C	60
	Elongation	72	30 s / kb	72	60 s / kb	72	60 s / kb
	Final elongation	72	600	72	60 s / kb	72	300
	Storage	8	∞	8	∞	8	∞

Samples were mixed with 10x DNA sample buffer after PCR and were analyzed using agarose gel electrophoresis, according to 5.2.1.2.

5.2.1.2. Agarose gel electrophoresis

After PCR, the PCR products were analyzed on an agarose gel. According to the gel size, approximately 2-6 µl Roti® GelStain (Carl Roth GmbH, Karlsruhe, Germany) were added to a 0.8% agarose gel. After addition of 1 µl 10x DNA sample buffer to 9 µl DNA sample, samples were loaded on the gel. Gels were run at 100 V for 60 min and inspected under UV light after the run. By comparing the DNA fragments to a DNA standard, the success of the respective PCR reaction or DNA digest was analyzed. After successful PCR, the remaining PCR product that was not loaded on the gel was purified using a PCR purification kit. In the presence of several bands visible on the agarose gel, the respective fragment of desired size was cut out using a scalpel and purified using a gel extraction kit according to the manual of the manufacturer.

5.2.1.3. DNA restriction

To create suitable 3' overhangs for ligation, PCR products and plasmids were digested using restriction enzymes. For plasmid digestion, it was necessary to add Antarctic Phosphatase for dephosphorylation of the 5' ends of the plasmid DNA and prevention of plasmid religation. The pipetting scheme for DNA restriction of PCR products and plasmid DNA is shown in table 6.

Table 6: Pipetting scheme for restriction digest of PCR products and vector DNA.

	PCR products	Plasmid DNA
Template DNA	x µl (ca. 1 µg)	x µl (ca 2 µg)
Restriction enzyme I	1 µl	1 µl
Restriction enzyme II	1 µl	1 µl
10x NEB buffer*	5 µl	5 µl
Antarctic Phosphatase	-	1 µl
10x buffer for Antarctic Phosphatase	-	5 µl
ddH_2O	to 50 µl	to 50 µl

* CutSmart buffer was used for all high fidelity (HF) restriction enzymes. For standard enzymes (no HF) buffers were used as proposed by the manual of the manufacturer.

DNA was digested for 2 h at 37 °C, followed by enzyme inactivation for 20 min at 65 °C. After digestion, the DNA samples were purified using a PCR purification kit according to the manual of the manufacturer. DNA concentrations were determined using a NanoDrop 1000 (VWR, Darmstadt, Germany). The UV absorption was measured at 600 nm and it was assumed that

a UV absorption of 1 equals 50 ng/µl double stranded DNA. DNA concentrations were then calculated accordingly.

5.2.1.4. Ligation

After restriction digest, PCR products and vectors were ligated using T4 DNA ligase. An insert:vector ratio of 3:1 was used for ligation reactions. The pipetting scheme for ligation reactions is shown in table 7.

Table 7: Pipetting scheme for ligation reactions using T4 DNA ligase.

	Ligase reaction
Insert	6 µl
Vector	2 µl
T4 DNA ligase	1 µl
10x buffer for ligase	1 µl

Ligation reactions were incubated for 12-16 h at 16 °C, followed by inactivation of T4 DNA ligase for 20 min at 65 °C. Ligated plasmids were immediately transformed into competent *E. coli* cells (chapter 5.4).

Analysis of cloning success was done by DNA sequencing. Therefore, circular DNA was extracted from *E. coli* cells (XL1blue or TOP10) by a plasmid MiniPrep Kit according to the protocol of the manufacturer. Samples were sequenced either inhouse at the Genome Analytics Facility (GMAK) at HZI (Braunschweig, Germany) or by MWG Eurofins Genomics (Ebersberg, Germany) according to the respective protocols. *ClustalOmega* (Sievers *et al.*, 2011) was used to analyze sequencing results by pairwise sequence alignments with insert DNA.

5.2.2. Sequence-and-ligation-independent cloning

Besides traditional cloning using restriction enzymes, sequence-and-ligation-independent cloning (SLIC) was used in this study. Therefore, oligonucleotides for PCR of the respective inserts were designed with overhangs complementary to the respective plasmid DNA used for cloning. This ensures annealing of single-stranded overhangs of PCR product and vector. For amplification of insert DNA, Phusion Flash PCR Mastermix (ThermoFisher Scientific, Waltham, Massachusetts, USA) was used. Cloning was performed as indicated in the respective manuals of the manufacturers and according to the protocols established at DPF (Dortmund, Germany). A pipetting scheme for SLIC-PCR reactions is shown in table 8. The standard PCR program for PCR using Phusion Flash Mastermix is summarized in table 9.

Table 8: Pipetting scheme for PCR reactions using Phusion Flash Mastermix for SLIC.

	Volume	final concentration
2x Phusion Flash Mastermix	15 µl	1x
Template DNA	2 µl	10 ng
Primer fwd	2 µl	0.5 µM
Primer rev	2 µl	0.5 µM
5 M betaine	6 µl	1 M
ddH_2O	3 µl	

Table 9: Standard PCR program used for SLIC in this study.

	Step	T (°C)	t (s)
	Initial denaturation	98	120
	Denaturation	98	10
30x	Annealing	T_m – 5 °C	10
	Elongation	72	45 s / kbp
	Final elongation	72	420
	Storage	8	∞

Immediately after PCR, a DpnI digest was performed to remove excessive template DNA. 1 µl DpnI and 3.4 µl 10x CutSmart buffer were added to the 30 µl PCR reaction and incubated for 1 h at 37 °C. After addition of 15.6 µl sterile ddH_2O, reactions were purified using a PCR purification kit according to the manual of the manufacturer. A small amount of digested PCR product was run on a 0.8% test agarose gel to check the success of the respective PCR reactions. Plasmids were digested by KpnI and HindIII for 2 h at 37 °C in CutSmart buffer. To create single-stranded overhangs, vector and insert DNA were treated with T4 DNA polymerase (SLIC mix 1) individually without adding dNTPs. A vector:insert ratio of 1:1 was used. The pipetting scheme for SLIC mix 1 is shown in table 10.

Table 10: Pipetting scheme for SLIC mix 1.

Component	Volume	Stock concentration	Final concentration
T4 DNA polymerase	3.3 µl	0.03 U/µl	0.1 U/reaction
NEB buffer 2	2 µl	10x	1x
ddH_2O sterile	14.7 µl – x		
vector / insert	x µl		25 ng vector

Each reaction was incubated for 10 min at 22 °C and stopped by addition of 2 µl 10 mM dCTP, which blocks the exonuclease activity of T4 DNA polymerase. For the recombination reaction (SLIC mix 2), insert and vector DNA with single-stranded complementary overhangs were

incubated with recombinase A to ensure ligation of insert and plasmid. The pipetting scheme for the recombination reaction is shown in table 11.

Table 11: Pipetting scheme for SLIC Mix 2 (recombination reaction) used for SLIC in this study.

	Volume	Stock concentration	Final concentration
T4 DNA ligase buffer	2 µl	10x	1x
RecA	1 µl	40 ng/µl	2 ng/µl
Vector SLIC Mix 1	8.5 µl		
PCR SLIC Mix 1	8.5 µl		

SLIC Mix 2 reactions were incubated for maximal 30 min at 37 °C and immediately transformed into highly competent *E. coli* Omnimax cells. Positive clones were identified by colony PCR and DNA sequencing.

5.2.3. Site-directed mutagenesis / Surface entropy reduction

To introduce specific changes in the DNA sequence, a slightly modified protocol of QuickChange® II site-directed mutagenesis (Stratagene) was used. In this study, site-directed mutagenesis was used to delete frameshift mutations, to introduce a mutation for inactivation of a protein and for generation of surface entropy reduction (SER) mutants. For this purpose, specific oligonucleotides were created, which contain the desired mutation in the middle of the oligonucleotide sequence. Forward and reverse primer have to be complementary to each other. The PCR pipetting scheme for site-directed mutagenesis using KAPAHifi™ DNA polymerase and the PCR protocol are summarized in table 4 and table 5. 16 cycles of denaturation, annealing and extension were used for mutation of a single amino acid, 18 cycles were used for multiple amino acid deletions or insertions.

5.3. Preparation of chemically competent *E. coli* cells

Chemically competent *E. coli* cells were used for molecular biology experiments and protein expression. Rubidium competent *E. coli* cells were prepared by Claudia Hanko (HZI, Braunschweig, Germany) according to established standard protocols. Cells were flash-frozen in liquid nitrogen and stored at -80 °C.

5.4. Heat shock transformation of *E. coli* cells

For plasmid preparation and protein production, plasmids were transformed into competent *E. coli* cells by heat shock transformation. 50 µl competent cells were mixed with 100 ng plasmid DNA and incubated on ice for 30 min. Heat shock was performed for 1 min at 42 °C and cells were incubated on ice for further 5 min. 1 ml pre-warmed LB medium was added and

the mixture was incubated for 45-60 min under shaking at 37 °C. 100 µl were plated on LB-agar plates containing respective antibiotics and stored in a 37 °C incubator overnight.

5.5. Test expression

To determine suitable conditions for protein expression, a test expression was performed for most constructs. Several *E. coli* strains (*E. coli* BL21 CodonPlus(DE3)RIL, *E. coli* Rosetta2 (DE3)), as well as different media (LB, TB, autoinduction) in combination with various expression temperatures were tested. Plasmids were transformed in competent *E. coli* expression strains for protein production. A single colony was inoculated in 50 ml LB medium containing respective antibiotics and was incubated over night at 37 °C in the shaking incubator. This overnight culture was used to inoculate a 100 ml culture on the next day to an OD_{600} ~ 0.05. Cells were incubated at 37 °C in a shaking incubator. When OD_{600} reached 0.8, 1 ml culture (sample before induction) was pelleted. Protein expression was induced in the rest of the cell culture by addition of 1 mM IPTG. The cells were incubated over night at the respective expression temperature. At different time points after induction (1 h, 3 h, overnight) OD_{600} was measured and samples were taken to analyze protein production on an SDS PAGE. The samples after induction were normalized to the respective OD_{600} before induction to analyze the time-dependent protein expression.

For cell lysis, the samples collected 1 h and 3 h after induction were treated with BugBuster® (Merck, Darmstadt, Germany) according to the protocol of the manufacturer. The samples collected after overnight protein expression were not treated with BugBuster but were lysed by sonication. Therefore, 10 ml samples of each culture were pelleted. The cell pellets were resuspended in 1-2 ml buffer (usually 20 mM Tris/HCl pH 8.0, 300 mM NaCl) and disrupted by sonication (2x 10 sec pulse with 10 sec break between each pulse), followed by 30 min centrifugation at 13.000 rpm. Samples of supernatant and pellet were stored for SDS PAGE. For test expression of proteins fused to a His_6-tag, 25 µl Ni-NTA superflow beads (Qiagen, Hilden, Germany) were added to the supernatant fraction. The mixture was incubated at 4 °C and washed several times by alternating steps of centrifugation (1.500 rpm, 1 min, 4 °C) and resuspension of the beads in 25-30 µl buffer. After the final washing step, the beads were resuspended in 25 µl buffer and a sample was retained for SDS PAGE. Various samples of supernatant, pellet and beads and respective time points were then loaded on an SDS-PAGE to analyze the soluble production of proteins.

5.6. Protein production

5.6.1. Protein expression in *Bacillus megaterium*

Some TcdB constructs (Strep-TcdB543-2366-His_6 and Strep-TcdB543-2366(C698A)-His_6) were expressed in *Bacillus megaterium*. Transformation of both constructs in *Bacillus megaterium* protoplasts was done by Prof. Dr. Ralf Gerhard (MHH, Hannover, Germany). Cells were plated on LB-agar including tetracycline. A single colony was used to inoculate overnight cultures (50 ml LB medium + 5 µg/µl tetracycline). For glycerol stock preparation, 500 µl overnight culture were mixed with 500 µl glycerol. The mixture was immediately fresh frozen in liquid nitrogen and stored at -80 °C. These glycerol stocks were used for inoculation of overnight cultures during the rest of this study. Therefore, a small amount of the glycerol stock was added to 50 ml LB medium including tetracycline using a pipette tip. The culture was incubated over night at 37 °C under vigorous shaking and this overnight culture was used for inoculation of 1 L cultures the next day. 8x 1 L LB medium including tetracycline was inoculated with the overnight culture to OD_{600} ~ 0.05. The large cultures were incubated at 175 rpm at 37 °C. When cells reached an optical density OD_{600} ~ 0.4, protein expression was induced by addition of 0.5% α-D-xylose. After shaking at 175 rpm for additional four hours, cells were harvested at 4 °C, 7000 rpm for 30 min. The cell pellets were flash frozen in liquid nitrogen and stored at -80 °C until further usage.

5.6.2. Protein expression in *E. coli*

5.6.2.1. Standard expression (T7 promoter)

In this study, genes were cloned in an expression vector behind a T7 promoter. The encoded protein is only expressed in the presence of T7 polymerase. Therefore, plasmids were transformed in *E. coli* strains, which contain T7 polymerase. This polymerase is controlled by the lac promoter. In the absence of lactose or presence of glucose (catabolite repression), the lac repressor is bound to the lac promoter and prevents expression of the recombinant protein. By adding lactose or the lactose analogue IPTG to the medium, protein expression is induced, as lactose induces the dissociation of the lac repressor (Studier *et al.*, 1990).
For protein production, expression constructs were freshly transformed in competent *E. coli* cells. Overnight cultures (LB medium including respective antibiotics) were inoculated with a single colony and grown over night at 37 °C under vigorous shaking. Large cultures (1 L medium containing respective antibiotics) were inoculated with overnight cultures to an OD_{600} of 0.05. The cultures were grown to an optical density OD_{600} of 0.8 - 1.0 at 37 °C under vigorous shaking. Protein expression was induced by addition of 1 mM isopropyl-thio-β-D-galactopyranosid (IPTG). Cells were incubated at 20 °C overnight and harvested by

centrifugation at 4 °C and 7.000 rpm for 20 min. Pellets were flash-frozen in liquid nitrogen and stored at -80 °C until further usage.
Expression conditions for proteins that were expressed and purified in this study are summarized in table 39 in the supplement.

5.6.2.2. Expression of selenomethionine-labeled proteins

Selenomethionine (SeMet)-labeled proteins were used for structure determination by SeMet-SAD for some proteins in this thesis. To express SeMet-labeled protein, the respective expression plasmid encoding the protein was freshly transformed into chemically competent *E. coli* cells. A single colony was used to inoculate an overnight culture (50 ml LB medium, containing respective antibiotics), which was incubated over night at 37 °C under vigorous shaking. The overnight cultures were pelleted the next day by centrifugation at 4.500 rpm, 4 °C for 30 min. The cell pellet was washed with 50 ml M9 medium two times and was finally resuspended in 40 ml M9 minimal medium. Large cultures (1 L M9 minimal medium, containing antibiotics) were inoculated with 20 ml of the washed overnight cultures and incubated at 37 °C in a shaking incubator. When the cells reached an OD_{600} ~ 0.5, amino acids were added (100 mg L-lysine, 100 mg L-threonine, 100 mg L-phenylalanine, 50 mg L-isoleucine, 50 mg L-leucine, 50 mg L-valine per liter medium). After another 30 min incubation at 37 °C, 60 mg selenomethionine was added per liter cell culture. Protein expression was induced by adding the respective amount of IPTG that was essential for native protein expression. Expression of SeMet-labeled protein was performed analogous to native protein expression. The cells were harvested at 4 °C, 7.000 rpm for 30 min, pellets were flash-frozen in liquid nitrogen and stored at -80 °C until further usage.

5.6.2.3. Autoinduction

In a T7 expression system, protein expression is typically induced by addition of IPTG. Using a defined autoinduction medium, no addition of IPTG is needed. One can control protein expression by adding a defined amount of glucose and lactose to ensure protein expression during the late log phase. Glucose will ensure the binding of the lac repressor to the lac promoter. Lactose will then ensure the removal of the repressor and induce protein expression. In the presence of both glucose and lactose in the medium, bacteria will first use all glucose for their metabolism because lactose is not the preferred carbohydrate source for *E. coli*. The protein expression will hence be turned on when all glucose in the medium was consumed by the bacteria. Therefore, one advantage of autoinduction is that there is no need to monitor the progress of protein expression and manual induction of protein expression by adding IPTG is not necessary (Studier, 2005).

For autoinduction, the expression plasmid was transformed in the respective *E. coli* cells. A single colony was used to inoculate an overnight culture (50 ml TB medium containing respective antibiotics, 1% glucose and 2 mM $MgSO_4$). The overnight cultures were grown overnight at 37 °C under shaking. Large cultures (1 L TB medium containing respective antibiotics, 0.01% lactose and 2 mM $MgSO_4$) were inoculated with the overnight cultures to an OD_{600} of 0.5. Cells were incubated at 37 °C in a shaking incubator for 4 h. Temperature was then decreased to 25 °C and cells were incubated for further 20 - 24 h. Cells were harvested by centrifugation for 25 min at 4 °C and 6.500 rpm. Cell pellets were flash-frozen in liquid nitrogen and stored at -80 °C until further usage.

5.7. Protein purification

5.7.1. Cell disruption

The relevant amount of cell pellet stored at -80 °C was thawed on ice. The pellet was resuspended in the appropriate volume of resuspension buffer. DNase A (1 mg/ml stock) and protease inhibitor (cOmplete™ Mini Protease Inhibitor Cocktail, Roche, Mannheim, Germany) were added prior to cell disruption. The cell pellet was homogenized by stirring on a magnetic stirrer at 4 °C. Cells were disrupted by 30 min sonication (1 s pulse and 10 s break between each pulse) using a Bandelin Sonopuls sonicator (BANDELIN electronic GmbH & Co. KG, Berlin, Germany). After sonication, cell debris and insoluble proteins were removed from the soluble protein fraction by centrifugation at 4 °C and 16.000 rpm for 1 h. Samples for SDS-PAGE were collected from the supernatant and pellet fraction after centrifugation to analyze the success of soluble protein expression.

5.7.2. Ni(II) Immobilized metal ion affinity chromatography

Ni(II)-Immobilized metal ion affinity chromatography (Ni(II)-IMAC) was performed using a 5 mL *HisTrap™ HP* column attached to an *ÄktaPurifier* FPLC system, which had been washed with degassed ddH_2O and the respective buffers. The column was also washed with ddH_2O and equilibrated with buffer A. The soluble protein fraction was loaded onto the column using a sample pump with a flow of 1 - 3 ml/min and a maximal pressure of 0.3 MPa. Unbound proteins were washed off with buffer A and unspecifically bound proteins were eluted by washing the column with 3% buffer B (15 mM imidazole). A linear gradient from 3 - 100% buffer B (15 - 500 mM imidazole) over ca. 15 CV (column volumes) was used for protein elution. 2 mL fractions were collected and particular samples were analyzed by SDS-PAGE.

5.7.3. Further affinity chromatography methods

Proteins carrying a His_6-MBP tag were purified using MBP-affinity chromatography. The procedure was performed analogous to Ni(II)-IMAC (chapter 5.7.2). The protein solution was loaded onto 1 – 3 MBPTrap™ HP 5 mL columns in series and elution was achieved by washing the column with 100% buffer B (10 mM maltose).
Proteins with a Strep-tag were purified using Strep-affinity chromatography analogous to MBP-affinity chromatography. Protein solution was loaded on a 10 mL column packed with Strep-Tactin® Superflow® and the target protein was eluted by 100% buffer B containing 5 mM D-desthiobiotin.

5.7.4. Dialysis and tag cleavage

After affinity chromatography, fractions containing the target protein were pooled. To cleave off the affinity tag, the appropriate amount of the respective protease was added. 1 mg Sumo protease was used to cleave 100 mg target protein, TEV- and S3C-protease were used in a 1:30 molar protein:protease ratio. The protein solution was dialyzed over night at 4 °C under continuous stirring against 2 L dialysis buffer using SnakeSkin™ dialysis membrane with a 3500 MWCO (ThermoFisher Scientific, Waltham, Massachusetts, USA).

5.7.5. Reverse Ni(II) immobilized metal ion affinity chromatography

After tag cleavage and dialysis overnight, precipitated protein was separated from the soluble fraction by centrifugation at 4.500 rpm 4 °C for 30 min. The supernatant was loaded on a 5 mL *HisTrap™ HP* column after centrifugation and another Ni(II)-IMAC run was performed according to 5.7.2 to separate the cleaved protein from the tag, protease und uncleaved protein. Cleaved protein should not be able to bind to the column anymore and should hence elute in the flowthrough. The cleaved His_6-tag, the His-tagged protease and uncleaved protein should instead bind to the column and elute during a linear gradient to 100% buffer B containing 500 mM imidazole. Protein containing fractions were collected and analyzed by SDS-PAGE. Depending on the purity of the target protein, it was subjected to another step of chromatography (e.g. ion exchange chromatography) to yield higher purity or was directly used for final size exclusion chromatography.

5.7.6. Ion exchange chromatography (IEC)

Depending on the pI value of the target protein, anion or cation exchange chromatography was performed. Before IEC, the respective buffer was exchanged to a low-salt buffer (containing maximal 20 mM NaCl). The protein solution was loaded onto a 5 mL ion exchange column, which was equilibrated with buffer A. Unbound protein was washed off with buffer A and a

linear gradient from 0 - 100% buffer B (0 - 1 M NaCl) over 150 ml was applied for protein elution. 2 mL fractions were collected and protein-containing samples were analyzed by SDS-PAGE. Fractions containing the target protein were pooled and concentrated prior to final size exclusion chromatography.

5.7.7. Size exclusion chromatography

Before final size exclusion chromatography, the target protein was concentrated using Vivaspin® concentrators (Sartorius AG, Göttingen, Germany) with an appropriate MWCO according to the size of the protein (chapter 5.7.10). Size exclusion chromatography columns were washed with one column volume degassed water and equilibrated with at least one column volume size exclusion chromatography buffer. The concentrated protein was then injected manually onto the size exclusion chromatography column using a sample loop. The size exclusion chromatography run was performed using a flow and maximal pressure recommended by the column manufacturer for the respective column. Samples of eluted proteins were analyzed using SDS-PAGE, fractions with pure target protein were pooled and concentrated. The protein was either directly used for crystallization screens or flash frozen in liquid nitrogen and stored at -80 °C until further usage.

5.7.8. Purification protocols

Unless stated differentially, all proteins used in this study were purified using a 3- or 4-step purification protocol. A summary of all proteins purified in this study with the according chromatography steps and final buffer is shown in table 12.

Table 12: Summary of all proteins purified in this study.

Protein (Construct)	Purification Protocol	Final buffer
TcdB 543-2366 (C698A) (pHIS1522-TcdB 543-2366 (C698A))	Strep-affinity chromatography no tag cleavage size exclusion chromatography	20 mM Tris/HCl pH 8.0, 300 mM NaCl
TcdB800-1834 (pOPINM-TcdB800-1834)	Ni(II)-IMAC MBP-affinity chromatography tag cleavage by S3C protease reverse Ni(II)-IMAC IEX size exclusion chromatography	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 10% glycerol
CD1363 (pOPINM-CD1363)	MBP-affinity chromatography tag cleavage by S3C-protease reverse Ni(II)-IMAC size exclusion chromatography	20 mM HEPES pH 7.5, 100 mM NaCl, 3 mM β-Me
CD1363 SeMet (pOPINM-CD1363)	MBP-affinity chromatography tag cleavage by S3C-protease reverse Ni(II)-IMAC size exclusion chromatography	20 mM HEPES pH 7.5, 100 mM NaCl, 3 mM DTT

CD1364 (pOPINM-CD1364)	MBP-affinity chromatography tag cleavage by S3C-protease reverse Ni(II)-IMAC size exclusion chromatography	20 mM HEPES pH 7.5, 100 mM NaCl, 3 mM β-Me
CD1219 (pET-Sumo-CD1219)	Ni(II)-IMAC tag cleavage by Sumo protease reverse Ni(II)-IMAC IEX size exclusion chromatography	20 mM Tris/HCl pH 8.0, 150 mM NaCl
CD1219 SeMet (pET-Sumo-CD1219)	Ni(II)-IMAC tag cleavage by Sumo protease reverse Ni(II)-IMAC IEX size exclusion chromatography	20 mM Tris/HCl pH 8.0, 150 mM NaCl, 2 mM DTT
CD1823 (p10$-CD1823)	Ni(II)-IMAC tag cleavage by Sumo protease reverse Ni(II)-IMAC size exclusion chromatography	20 mM Tris/HCl pH 8.0, 150 mM NaCl
CD2589 (pET-Sumo-CD2589)	Ni(II)-IMAC tag cleavage by Sumo protease reverse Ni(II)-IMAC size exclusion chromatography	20 mM Tris/HCl pH 8.0, 150 mM NaCl
CD2752 (pET-Sumo-CD2752)	Ni(II)-IMAC tag cleavage by Sumo protease reverse Ni(II)-IMAC size exclusion chromatography	20 mM Tris/HCl pH 8.0, 150 mM NaCl, 2 mM DTT

5.7.9. SDS-PAGE

The success of soluble protein expression and purification was analyzed using Sodium-Dodecyl-Sulfate Gel Electrophoresis (SDS-PAGE). Gels were prepared using standard protocols (Laemmli, 1970) and consisted of a 5% polyacrylamide stacking gel and 10%, 15% or 18% polyacrylamide resolving gels, respectively. Protein samples were diluted with buffer to an appropriate protein amount, 8X protein sample buffer was added and the samples were denatured at 95 °C for 5 min. After short centrifugation at 13.000 rpm, samples were run on a gel at 150 V for ca. 40 - 60 min in running buffer. Afterwards the gel was stained with InstantBlue™ (Expedeon, Cambridgeshire, UK) staining solution for 15 min on a shaker and was washed with water afterwards.

5.7.10. Concentration of protein samples

Protein samples have to be concentrated before size exclusion chromatography and crystallization. Therefore, Vivaspin® concentrators (Sartorius AG, Göttingen, Germany) with an appropriate MWCO depending on the protein size were used in this study. The concentrator membrane was washed with buffer prior to use by centrifugation at 4.000 rpm for 5 min. After buffer removal, the protein sample was loaded in the concentrator. The protein sample was

concentrated by centrifugation at 4 °C and 4.000 rpm until the desired protein volume was reached.

5.7.11. Determination of protein concentration

Protein concentrations were determined using a NanoDrop 1000 (VWR, Darmstadt, Germany) spectrophotometer. The theoretical molecular weight und extinction coefficient of the protein (table 13) were calculated based on the amino acid sequence using EXPASY ProtParam (Gasteiger *et al.*, 2003). The absorption of the protein sample was measured at 280 nm. The protein concentration was calculated using the Lambert-Beer's law (equation 5).

$$A = c \cdot d \cdot \varepsilon_{280} \quad (5)$$

In this equation, *A* is the absorption of the protein solution at 280 nm, *c* is protein concentration in M, *d* is path length in cm and ε_{280} is the extinction coefficient at 280 nm in $M^{-1}cm^{-1}$.

Table 13: Overview of theoretical molecular weight and extinction coefficient of all proteins purified in this study.

Protein	Molecular weight (kDa)	Extinction coefficient ($M^{-1}cm^{-1}$)
Strep-TcdB543-2366-His_6	210.1	225430
TcdB800-1834	116.3	92600
CD1363	39.2	20525
CD1364	16.0	23950
CD1067	44.8	8900
CD1219	39.5	23840
CD1823	29.2	27850
CD2589	34.3	20400
CD2752	25.6	21890

5.8. Protein crystallization and structure determination

5.8.1. Initial crystallization screens

Screening for crystallization conditions was performed using the sitting drop vapor-diffusion technique in a 96 well INTELLI-PLATE® (Art Robbins Instruments, Sunnyvale, USA). Plates were set up using the *Honeybee 961* (Zinsser Analytic GmbH, Frankfurt, Germany) dispensing robot. Drops consisted of 0.2 µl protein solution and 0.2 µl reservoir solution, the reservoir volume was 70 µl. Different commercially available screens (chapter 5.1.8) in combination with various protein concentrations (20 mg/ml, 15 mg/ml, 10 mg/ml, 5 mg/ml) were used for initial screening. All crystallization plates were set up at 20 °C. Crystallization plates were stored in

an automated imaging system (Formulatrix) and were imaged regularly. Images were inspected and analyzed manually via the Rockmaker software (Formulatrix).

5.8.2. Crystal optimization

Some initial crystallization hits were optimized using grid or random screens. Optimization of crystallization conditions is necessary when crystals are e.g. too small for fishing or resolution is insufficient for data collection. In a grid screen, one crystallization condition is optimized by changing one parameter in each row or column of the screen, respectively, e.g. protein concentration, buffer pH, concentration of salt or precipitant, etc. The Formulator system (Formulatrix) was used to prepare random screens. Thereby, ingredients of several initial crystallization hits are randomly combined to find new crystallization conditions, which are not covered by using commercially available screens. Plates for crystal optimization were set up according to 5.8.1 using the *Honeybee* robot.

5.8.3. Data collection

Unless stated differently, crystals were cryoprotected using 10% 2,3-butanediol added to the mother liquor. The crystals were fished with nylon loops and were flash-frozen in liquid nitrogen. Screening for crystal diffraction was performed inhouse (HZI, Braunschweig, Germany) at 100 K. Test images were collected using a wavelength of λ = 1.54 Å (Cu-K_α edge) and a Rigaku Saturn 944+ CCD detector. High-resolution datasets, as well as anomalous data, were collected at SLS (Swiss Light Source, Villigen, Switzerland; Müller *et al.*, 2012), DESY (Deutsches Elektronen Synchrotron, Hamburg, Germany; Burkhardt *et al.*, 2016) and BESSY (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung m. b. H., Berlin, Germany; Müller *et al.*, 2015). Datasets that were used for structure determination in this study are listed in table 14.

Table 14: Data collection strategies for protein crystals used in this study.

Protein	Beamline	Detector	Wavelength λ (Å)	Images	Oscillation (°)
native CD1219	PXII, SLS	PILATUS 6M	1.000	3600	0.1
SeMet-CD1219	PXII, SLS	PILATUS 6M	0.978	7200	0.1
native CD1823	P11, DESY	PILATUS 6M	1.033	3600	0.1
SeMet-CD1363	PXIII, SLS	PILATUS 6M	0.979	7200	0.1
native CD1364	BL14.1, BESSY	PILATUS 6M	0.918	3600	0.1

5.8.4. Data processing and structure determination

Data were integrated using *XDS* (Kabsch, 2010) and scaled using *Aimless* (Evans, 2006). Molecular Replacement was performed using *Phaser* from the *Phenix* (Adams *et al.*, 2011) software. SeMet-SAD data were analyzed using the *ShelX* software (Sheldrick, 2008) and *Autosol* (Adams *et al.*, 2010) for data processing. After structure solution, the model was built manually and refined in iterative cycles. *COOT* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010) was used for model building and *phenix.refine* (Afonine *et al.*, 2012) for refinement. Final structures were validated using *MolProbity* (Chen *et al.*, 2009) and figures were prepared using *PyMOL* (DeLano, 2002) and *Chimera* (Pettersen *et al.*, 2004).

5.8.5. Limited proteolysis

To find more stable protein fragments for protein crystallization, limited proteolysis experiments were performed using Proti-Ace™ and Proti-Ace™ 2 Kit (Hampton Research). The protein solution was diluted to 10 mg/ml in 20 mM Tris/HCl pH 8.0, 150 mM NaCl. Proteases were diluted to 0.01 mg/ml with 10 mM HEPES pH 7.5, 500 mM NaCl. 10 µl protein was mixed with 10 µl protease and incubated at 37 °C. After 1 h, 2 h and 16 h, protein digestion was stopped by addition of 5 µl 8X SDS sample buffer to 8 µl digested protein sample. 10 µl were run on a 15% SDS-PAGE for analysis of protein digestion by the individual proteases.

5.9. Small angle X-ray scattering

5.9.1. Principles of Small angle X-ray scattering

Small angle X-ray scattering (SAXS) is an analytical method that is used for the structural analysis of proteins in solution at low resolution. It uses the same principle as X-ray crystallography: a protein sample is exposed to a focused X-ray beam, X-rays are diffracted by the protein sample and the scattered X-rays are recorded by a detector. In contrast to X-ray crystallography, the experiments are performed in solution and not in the crystalline state. Due to the random orientation of proteins in solution, structural information is lost during these experiments, which leads to a resolution of approximately 10 - 20 Å. Hence, one can get insight into the overall shape and the oligomeric state of the macromolecule in solution, but resolution is not high enough to determine the protein structure. SAXS is thus often used in combination with high-resolution techniques for structure determination, such as X-ray crystallography (Putnam *et al.*, 2013). As SAXS is a contrast method, it is essential to measure a buffer blank. This blank has to be subtracted from the protein data to yield the diffraction of the protein sample exclusively without buffer background (Putnam *et al.*, 2007).

The protein solution is hit by X-rays of a certain wavelength, usually between 1 - 1.5 Å and intensities of the scattered X-rays are recorded by the detector at a low angle respective to the

incident beam (θ = 0.1 - 10°). The scattered intensities are recorded as a function of the scattering angle, resulting in a scattering curve. The intensity is dependent on the scattered vector $|\vec{s}|$, also called momentum transfer. This momentum transfer is described by the wavelength λ and the angle θ, where 2θ is the angle between incident and scattered beam (Feigin & Svergun, 1987).

$$|\vec{s}| = \frac{4\pi \cdot \sin\theta}{\lambda} \tag{6}$$

Several important parameters can be directly derived from the scattering curve, enabling a fast characterization of the protein sample in regards of size, shape and oligomeric state of the protein. These parameters are e.g. the maximum particle diameter D_{max}, molecular mass of the protein (MM), radius of gyration R_g and the volume of the hydrated particle V_p. The so-called Guinier analysis of the experimental data is the most straightforward analysis to derive important parameters such as I_0 and R_g (Guinier, 1939). R_g is the radius of gyration and describes the overall size of the particles. I_0 is the intensity at $\theta = 0$. The intensity I_s of the scattered vector is described by the Guinier equation (equation 7) for globular proteins ($s \cdot R_g < 1.3$).

$$I_{\vec{s}} = I_0 \cdot e^{(-\frac{1}{3}R_g^2\vec{s}^2)} \tag{7}$$

By plotting ln(I_s) against s^2 in a Guinier plot, one directly gets information about R_g and I_0. R_g can be derived from the slope and I_0 from the y-axis intercept. The Guinier plot also gives an indication of the sample quality. A linear plot illustrates a monodisperse, globular folded sample (Guinier & Foumet, 1955). Non-linear plots indicate aggregation or an elongated shape of the protein (Fig. 8).

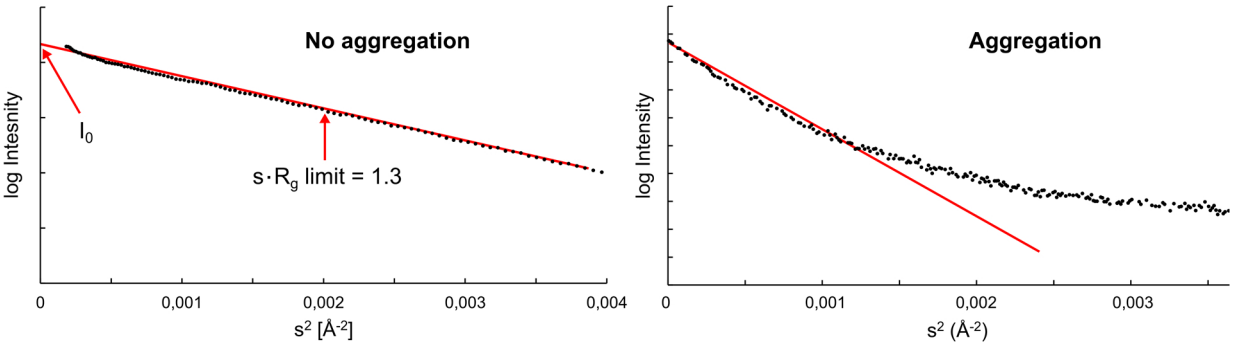

Figure 8: Guinier Plot of a non-aggregated (left) and an aggregated protein sample (right). A Guinier Plot is a non-linear plot of s^2 against logarithmic intensities. I_0 can be directly derived from the y-intercept of the plot, R_g can be derived from its slope. The shape of the Guinier-Plot gives an estimation of the sample quality, as a linear plot illustrates a high-quality and non-aggregated protein sample.

By comparison of I_0 and protein concentrations with a standard protein (e.g. BSA), one can obtain an estimation of the molecular weight of the measured protein (equation 8; Smilgies & Folta-Stogniew, 2015; Mylonas & Svergun, 2007).

$$MM_{Protein} = I_{0,Protein}/c_{Protein} \frac{MM_{Standard}}{I_{0,Standard}/c_{Standard}} \quad (8)$$

Another important parameter used for data processing is V_p, the excluded volume of the hydrated particle. It can be calculated by the Porod equation (Porod, 1982) with Q as Porod invariant (equation 9).

$$V_P = \frac{2\pi^2 I_0}{Q}, \quad Q = \int_0^\infty \vec{s}^2 I_{\vec{s}} d\vec{s} \quad (9)$$

The parameters R_g and I_0 can also be extracted by Fourier-transformation of the scattering intensities. This Fourier-transformation results in the distance-distribution function p(r) (Koch *et al.*, 2003; equation 10).

$$p(r) = \frac{1}{2}\pi^2 \int_0^\infty I_s \, \vec{s}\vec{r} \sin(\vec{s}\vec{r}) \, d\vec{s} \quad (10)$$

In this function, r is the distance between two electrons in the particle. The distance-distribution function $p(r)$ represents all pairwise distances in the protein in real space. The graphical shape of this function gives estimations about the protein shape (Fig. 9). Globular proteins usually have a bell-shaped distance-distribution function with a maximum at $D_{max}/2$. An *ab initio* model, as well as a theoretical scattering curve for this model can be calculated based on the distance-distribution function. This theoretical scattering curve is compared to the experimentally measured scattering curve. The theoretical model is then altered until theoretical and experimental scattering curves fit. One control parameter to evaluate the quality of this fit is the X-value (equation 11) (Grant *et al.*, 2015).

$$X^2 = \frac{1}{N} \sum_{j=1}^{N} (I_{exp,\vec{s}_j} - \frac{\eta I_{calc,\vec{s}_j}}{\sigma_{\vec{s}_j}})^2 \quad (11)$$

In this equation, N is the number of experimental points. $I_{exp,sj}$ and $I_{calc,sj}$ are the experimentally measured and theoretically calculated intensities, respectively. η is a scaling factor and σ_{sj} is the experimental error. A SAXS model of high quality is indicated by a X-value < 5.

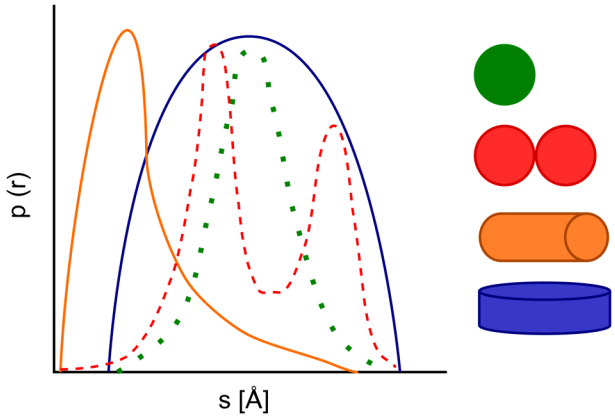

Figure 9: The distance distribution function *p*(r) is a Fourier-transformation of the scattered intensities of the SAXS sample. The shape of the function gives information about the shape of the measured protein.

5.9.2. Experiments and data evaluation

Protein samples for SAXS measurements were prepared by dialyzing the protein sample against SEC buffer. A dilution series (10 mg/ml, 5 mg/ml, 2.5 mg/ml, 1.25 mg/ml and 0.625 mg/ml) of each protein was prepared and the dialysis buffer was used as a blank for the measurements. Data were collected at BM29, ESRF (European Synchrotron Radiation Facility, Grenoble, France; Pernot *et al.*, 2013) using a PILATUS 1M detector, a sample detector distance of 2.867 m and a wavelength of λ = 0.9919.

The experimental data of the buffer blank were subtracted from the experimental protein data using *PRIMUS* software (Konarev *et al.* 2003). This software was also used for normalizing, averaging and merging the experimental data. The program GNOM (Svergun, 1992) was used for evaluation of experimental parameters, such as radius of gyration R_g, zero intensity I_0, maximum diameter D_{max} of the particle and the distance-distribution function $p(r)$. Using *DAMMIF* (Franke & Svergun, 2009), twenty runs of *ab initio* shape determination of the SAXS envelope were performed. A comparison of all models generated by *DAMMIF* was performed by *damsel*. This program also selected the most promising of all 20 models, determined and discarded possible outliers. An alignment of the selected models with the most promising one was conducted by *damsup*. *damaver* then averaged the models and computed a model map of the aligned and averaged models. The final SAXS model emerged after filtering the averaged model at a certain cut-off threshold by *damfilt*. Crysol (Svergun *et al.*, 1995) was used to superimpose the calculated SAXS envelope and the crystal structure of the protein. The SAXS envelope was created using *SITUS* (Wriggers, 2012) and illustrated with *Chimera* (Pettersen *et al.*, 2004) or *PyMOL* (DeLano, 2002).

5.10. CD spectroscopy

5.10.1. Theory of CD spectroscopy

In this study, CD (circular dichroism) spectroscopy was used to determine the secondary structure composition and quality of proteins in solution. Circular dichroism describes the phenomenon that chiral molecules can absorb left and right circularly polarized light to a different extent. In general, absorption of light is described by Lambert Beer's law (equation 12), with I_0 as intensity of the incoming light, I_1 intensity of the outcoming light, ε extinction coefficient, c concentration of the sample and d as pathlength (Bulheller *et al.*, 2007).

$$E = log_{10}\left(\frac{I_0}{I_1}\right) = \varepsilon \cdot c \cdot d \qquad (12)$$

As chiral molecules show different extinction coefficients for left and right handed circularly polarized light ($\varepsilon_L \neq \varepsilon_R$), $\Delta\varepsilon$ is recorded during CD spectroscopy ($\Delta\varepsilon = \varepsilon_L - \varepsilon_R$) as a function of the wavelength. $\Delta\varepsilon$ is then used to calculate the molar ellipticity (equation 13), which is monitored during CD spectroscopy.

$$\theta = const \cdot (\varepsilon_L - \varepsilon_R) \cdot c \cdot d \qquad (13)$$

The absorption of light is dependent on the geometry of the optically active molecule. By comparing the CD-spectrum of a protein in a specific wavelength region with exemplary spectra of secondary structure motifs, one can get an idea about the secondary structure composition of the protein. The amide region (250 - 170 nm) is typically influenced by the peptide bond region of a protein. Specific secondary structure motifs enhance the absorption of light in this wavelength region and show a characteristic CD spectrum. Typically, α-helical proteins show two minima at 220 and 208 nm, β-sheets show a minimum at 215 nm (Fig. 10; Bulheller *et al.*, 2007). Unfolded proteins usually give a minimum below 200 nm and a maximum around 210 - 220 nm. By measuring a CD spectrum of a protein sample, one can therefore directly get information of the proper folding of the protein.

5.10.2. Experimental procedure

Prior to spectrum collection, proteins were concentrated to 1 mg/ml in the respective size exclusion chromatography buffer and diluted with ddH_2O to 0.1 mg/ml. The buffer was also diluted 1:10 with buffer for a blank measurement. CD spectra were collected at Jasco J-815 CD spectrometer (Jasco, Gross-Umstadt, Germany) in 0.5 mm cuvettes (Hellma, Müllheim, Germany) from 190 – 300 nm, with 1 nm step resolution and 100 nm/min scanning speed. Ten spectra were measured for each protein and an averaged spectrum was calculated using Jasco Spec-Man-II software after buffer subtraction.

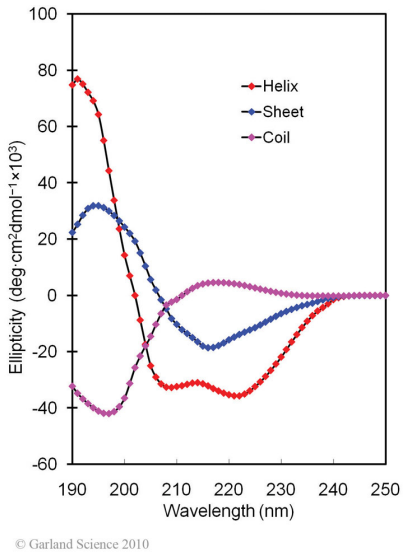

Figure 10: Exemplary CD spectra of secondary structure motifs. The CD spectrum of an α-helical protein is shown in red, the CD-spectrum of β-sheets in blue and that of random coils in magenta. Reproduced with permission from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC.

5.11. Thermal shift assay

To identify the melting temperature of a protein and to determine the optimal buffer for protein purification, a thermal shift assay can be performed. The melting point of a protein in the respective buffer is an indicator of protein stability and differs between used buffers. More stable proteins unfold at higher temperatures compared to unstable proteins. Proteins in a suitable buffer therefore show a higher melting temperature than proteins in a buffer that is not suitable for protein purification and crystallization. Protein denaturation by temperature is monitored by observing the fluorescence of a fluorescent dye, in this case SYPRO orange. This dye binds to hydrophobic patches of the protein. In the native folded state, these hydrophobic patches are hidden inside the protein. The dye is not bound to the protein and therefore no (or only low) fluorescence is observed. When the protein is unfolded upon temperature increase, hydrophobic patches are exposed. The dye binds to these hydrophobic patches, which results in an increase of fluorescence. Protein unfolding can therefore be directly observed by measuring the fluorescence as a function of temperature (Huynh & Partch, 2015).

Prior to the thermal shift assay, a concentration test has to be performed to find the optimal experimental parameters (protein concentration, SYPRO orange concentration) for this assay. The SYPRO orange stock (5000X, Invitrogen) was diluted to 500x, 100x and 10x with ddH_2O. The respective proteins were diluted to 2 mg/ml, 1 mg/ml and 0.5 mg/ml in the respective buffer. Different combinations of SYPRO orange and protein concentrations were tested in a 96 well plate (Multiplate™ 96-Well Unskirted PCR plate, BIO-RAD, Hercules, USA). For this, 5 µl of protein solution, 5 µl SYPRO orange solution and 40 µl of the respective buffer were

pipetted into a 96 well plate, the plate was sealed with Microseal®'B' Adhesive Seal (BIORAD) foil and the measurement was performed using a *CF*X96™ RT-PCR device (BIO-RAD).
After determining the best experimental parameters, the thermal shift assay was performed. Protein and SYPRO orange solution were diluted to the optimal concentrations, 5 µl of each solution were filled into 96-well plate and 40 µl buffer solution was added. Buffer screens were either prepared by Kevin Walkling (HZI, Braunschweig, Germany) or commercially available buffer or addivitive screens were used (RUBIC Additive screen, Molecular Dimensions). The plate was sealed with Microseal®'B' Adhesive Seal (BIORAD) foil and was heated in 1 °C steps from 4 °C to 95 °C in a *CF*X96™ RT-PCR device (BIO-RAD). The fluorescence was measured using an excitation wavelength of 485 nm and emission wavelength of 575 nm. Melting curves were plotted and melting temperature was calculated using *CF*X Manager™ software (BIO-RAD).

6. Toxins of *Clostridium difficile*

6.1. TcdB as one of the main virulence factors of *Clostridium difficile*

Clostridium difficile produces several virulence factors, among these are its two most prominent ones, the secreted protein toxins Toxin A (TcdA) and Toxin B (TcdB). The homologous toxins belong to the family of large clostridial toxins (LCTs). This family includes many other toxins, e.g. α-toxin from *Clostridium novyi* and lethal toxin from *Clostridium sordellii* (table 15; Schirmer & Aktories, 2004).

Table 15: Overview of toxins from the family of large clostridial toxins (LCTs).

Toxin	Molecular Weight (kDa)	Strain
Toxin A (TcdA)	308	*Clostridium difficile*
Toxin B (TcdB)	270	*Clostridium difficile*
Hemorrhagic toxin (TcsH)	300	*Clostridium sordellii*
Lethal toxin (TcsL)	270	*Clostridium sordellii*
α-toxin (Tcnα)	250	*Clostridium novyi*
Large cytotoxin (TpeL)	191	*Clostridium perfringens*

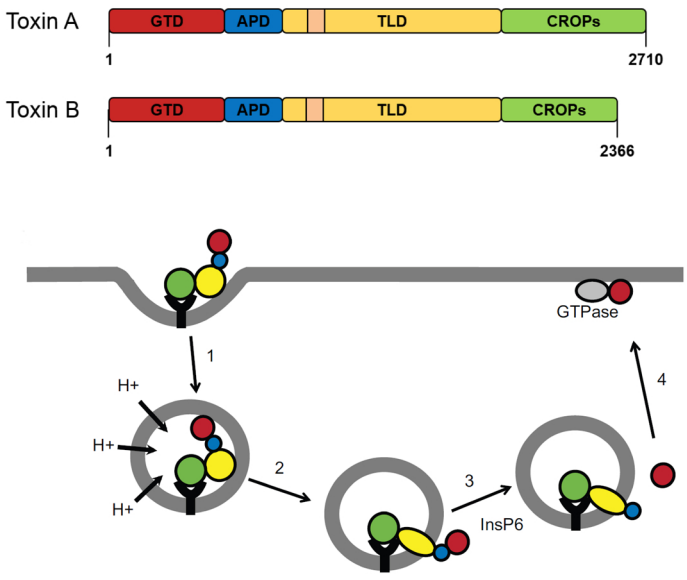

Figure 11: The homologous toxins TcdA and TcdB of *Clostridium difficile* show the same domain organization and mechanism of action. The top panel illustrates the domain architecture of the toxins. Respective domains are shown in different colors: glucosyltransferase domain (GTD) in red, autoprotease domain (APD) in blue, translocation domain (TLD) in yellow with its hydrophobic region in orange and the C-terminal receptor binding domain (CROPs domain) in green. A simplified schematic mechanism is shown in the bottom panel. After binding to host cell receptors, the toxin is taken up via clathrin-mediated endocytosis. Due to the acidic environment in the endosome, hydrophobic patches of the translocation domain are exposed and insert into the endosomal membrane to form a pore, through which the APD and GTD are transferred out into the cytosol. The APD cleaves off the GTD in the presence of $InsP_6$ and the free GTD can then glycosylate and inactivate host cell GTPases. The bottom part of the figure is adapted from Pruitt & Lacy, 2012.

TcdA and TcdB are high molecular weight toxins (308 kDa and 270 kDa, respectively). They share 49% sequence identity and the same structural organization, which is illustrated in Fig. 11. In general, they consist of four domains that mediate the toxin mechanism of action (Shen, *et al*., 2012; Chumbler, *et al*., 2016): after toxin binding to the respective receptors on the host cell surface, the toxin is internalized via clathrin-mediated endocytosis (Papatheodorou *et al*., 2010). After acidification of the endosome, a pore is formed. The glucosyltransferase domain is cleaved off by the autoproteolytic domain and is transferred into the cytosol, where it inactivates host cell GTPases by glucosylation (Pfeifer *et al*., 2003).

6.2. Domain architecture of TcdB

Toxin B of *Clostridium difficile* consists of four domains: an N-terminal glucosyltransferase domain, followed by an autoprotease domain, an intermediate domain and a C-terminal CROPs domain (Jank & Aktories, 2008). In the following chapter, the structure, function and the current state of research of the individual TcdB domains will be discussed in detail.

6.2.1. The C-terminal receptor binding domain

The C-terminal domain of TcdB (aa 1834-2366) is characterized by several repetitive elements, the CROPs (combined repetitive oligopeptides). TcdB contains 19 short (20-23 aa) and 4 long (30 aa) repeats that are arranged in a repetitive manner (Fig. 13C). These CROPs form cell wall binding motifs and are predicted to bind to receptors on the host cell surface due to structural homology to other receptor binding domains of homologous toxins (Orth *et al*., 2014). Nevertheless, it was recently shown by several studies that the CROPs domain is not the only receptor binding domain of TcdB. Truncation studies showed that a TcdB construct that lacks the CROPs domain can still enter host cells (Genisyuerek *et al*., 2011; Olling *et al*., 2011). In addition, other homologous toxins are known that completely lack a CROPs domain, e.g. TpeL from *Clostridium perfringens* (Schorch *et al*., 2014). Thus, there must be a second receptor-binding domain present in TcdB upstream to the CROPs domain. A minimal fragment essential for cellular uptake was identified with TcdB1-1493 (Manse & Baldwin, 2015).

Three receptors for this “additional” receptor binding domain of TcdB have been identified:

- Poliovirus receptor-like protein 3 (PVRL3; LaFrance *et al*., 2015): expressed on the surface of human colon epithelial cells
- Chondroitin sulfate proteoglycan 4 (CSPG4; Yuan *et al*., 2015): present in intestinal subepithelial myofibroblasts of mouse and human intestines

- Members of the Wnt receptor Frizzled protein family (Tao *et al.*, 2016): components of the surface epithelium of the human colon

Although it was shown that the CROPs domain is not essential for receptor binding, it has to play a major role during this first step of the toxin mechanism. One function of the CROPs domain could be to direct the toxin to the host cell by unspecific binding to the host cell surface, thereby directing the second receptor binding domain of TcdB to the specific receptors on the surface. Two antibodies (actoxumab, bezlotoxumab) that are directed against the CROPs domain of TcdA and TcdB, respectively, show a neutralizing effect of toxin cytotoxicity by blocking the binding of the toxin to host cell receptors (Yang *et al.*, 2015).

Crystal structures of TcdB CROPs fragments have been solved in complex with FAB domains of the neutralizing antibody bezlotoxumab (PDB-Code 4NP4; Orth *et al.*, 2014) and with a non-neutralizing antibody (PDB-Code 4NC2; Murase *et al.*, 2014). Cryo-EM and SAXS studies revealed the overall shape of the full-length TcdB CROPs domain. The TcdB CROPs domain shows a horseshoe-shaped architecture, but a crystal structure of the full-length CROPs domain of TcdB has not been determined yet (Albesa-Jové *et al.*, 2010; Pruitt *et al.*, 2010).

6.2.2. The intermediate translocation domain

After receptor binding, the toxin is internalized via clathrin-mediated endocytosis. Due to the acidic environment in the endosome, the intermediate domain (TcdB800-1834) undergoes conformational changes and exposes hydrophobic parts. These regions insert into the endosomal membrane to form a pore (Qa'Dan *et al.*, 2000). A low-resolution electron microscopy (EM)-envelope revealed the conformational changes of the translocation domain upon pH change during endocytosis. Upon pH decrease, it changes its conformation from a bilobed structure to a more elongated conformation (Pruitt *et al.*, 2010).

The intermediate domain of TcdB is responsible for pore formation and translocation of the enzymatically active part of the toxin into the host cell cytosol (Zhang *et al.*, 2013). It contains a hydrophobic region (TcdB980-1107) that consists mainly of hydrophobic residues, e.g. isoleucine, glycine, leucine, valine. Several residues of this hydrophobic region are involved in pore formation of TcdB (Genisyuerek *et al.*, 2011; Zhang *et al.*, 2014). Zhang *et al.* (2013) identified eight residues between amino acids 1035 and 1107 that are involved in TcdB pore formation by mutational studies.

Proteins that form membrane-inserting β-barrels typically contain a motif of alternating hydrophobic and hydrophilic parts. This motif is missing in the membrane-inserting part of TcdB. This indicates that the hydrophobic region of TcdB inserts into the endosomal membrane

via α-helical secondary structures and not as a β-barrel (Zhang *et al.*, 2014). Due to the high homology of TcdB and diphtheria toxin, it can be assumed that both toxins use a similar mechanism of membrane insertion, as both toxins also show the same pattern of hydropathy. Diphtheria toxin inserts into the endosomal membrane as a double hairpin that consists of four α-helices, which form two connected helical hairpins, called “double-dagger”-motif (Choe *et al.*, 1992; Wang *et al.*, 2009). Besides a similar toxin mechanism and a strikingly similar hydrophobicity pattern between TcdB and diphtheria toxin, the eight TcdB residues that are involved in pore formation (Zhang *et al.*, 2013) are conserved between TcdA and DT (Fig. 12). This points to the fact that both toxins use a similar way of pore-formation. It is hence possible that also TcdB of *Clostridium difficile* inserts into the endosomal membrane via helical secondary structure motifs similar to the double-dagger motif of diphtheria toxin.

Figure 12: Sequence alignment of the TcdB regions that are involved in pore formation with the respective regions of TcdA from *Clostridium difficile* and diphtheria toxin (DT) from *Corynebacterium diphtheria*. Amino acids that were shown to be involved in pore formation of TcdB (Zhang *et al.*, 2013) are marked with an orange asterisk. Conserved residues are highlighted by a black box. Residues with 70% similarity concerning physical-chemical properties are shown in bold letters.

A crystal structure of the TcdB intermediate domain was not known at the onset of this thesis and hence was a major goal at that time. Furthermore, the exact molecular changes that take place in the intermediate domain upon pH change have been not known. A crystal structure and further biochemical and biophysical characterization of this domain would be necessary to draw conclusions about the mechanism of membrane insertion and pore formation of TcdB.

6.2.3. The autoprotease domain

After pore formation, the biologically active glucosyltransferase and autoproteolytic domain of TcdB are transferred out of the endosome. In the presence of eukaryotic inositol hexakisphosphate ($InsP_6$) in the cytosol, the autoproteolytic domain cleaves off the glucosyltransferase domain (Egerer *et al.*, 2009).

The autoprotease domain of TcdB is homologuos to cysteine proteases of the MARTX toxin family (Shen *et al.*, 2011). The catalytic triad (Asp593, His653, Cys698) coordinates a zinc ion that was shown to be essential for the autoproteolytic activity of TcdB (Chumbler *et al.*, 2016). The binding site for $InsP_6$ is located distal to the active site of the TcdB APD. Both sites are separated from each other by the secondary structure motif “β-flap”, which blocks the active site (Fig. 13B). Upon $InsP_6$ binding, the “β-flap” undergoes a conformational change and the

active site gets accessible (Pruitt *et al.*, 2009). The APD is then activated and can cleave off the GTD.

The structure and mechanism of the TcdB APD are already well characterized due to the presence of crystal structures in complex with $InsP_6$ (PDB-ID 3PEE; Shen *et al.*, 2011) and with a peptide inhibitor (PDB-ID 3PA8; Puri *et al.*, 2010).

6.2.4. The N-terminal glucosyltransferase domain

After cleavage by the APD, the N-terminal glucosyltransferase domain of TcdB can glucosylate and hence inactive host cell Rho-family GTPases such as RhoA, Rac1 and Cdc42 (Busch *et al.*, 1998). The GTD (aa 1-543) of TcdB shows a typical glucosyltransferase type A fold with a central β-sheet and a common DXD motif (Asp286-Val287-Asp288; Fig. 13A). This motif is involved in binding the glucose-moiety of the cosubstrate UDP-glucose and coordination of a Mn^{2+} that is essential for catalytic activity of the GTD (Reinert *et al.*, 2005). It transfers the glucose moiety from the cosubstrate UDP-glucose to a threonine residue (Thr35/37) of host cell Rho-family GTPases, whereupon these are inactivated. This inactivation of GTPases affects several downstream signaling processes, leading to a rearrangement of the actin cytoskeleton, cell rounding and apoptosis (Brito *et al.*, 2002).

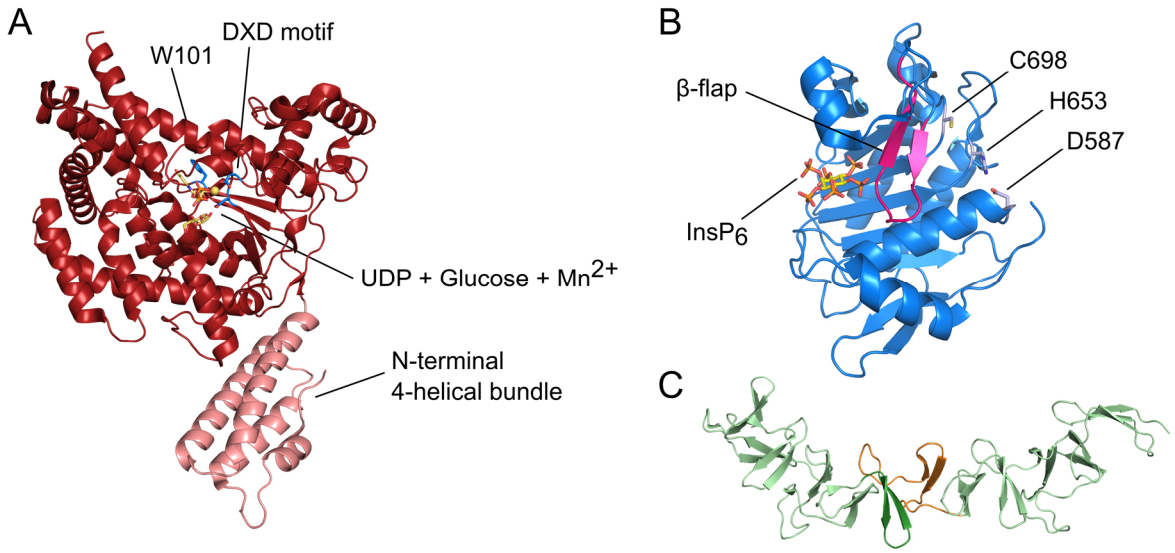

Figure 13: Crystal structures of individual TcdB domains in cartoon representation. (A) The crystal structure of the TcdB glucosyltransferase domain (TcdB1-543; PDB-ID 2BVL, Reinert *et al.*, 2005) is shown in dark red in complex with its cosubstrate UDP-glucose (shown as yellow sticks) and Mn^{2+} (shown as yellow sphere). Conserved residues (W101, DXD motif) are shown as blue sticks; the N-terminal 4-helical bundle is highlighted in light red. (B) The crystal structure of the autoprotease domain of TcdB (TcdB544-797; PDB-ID 3PEE, Shen *et al.*, 2011) is illustrated in blue with bound $InsP_6$ in yellow. The catalytic triad, which is essential for autoproteolytic activity of TcdB (C698, H653, D587), is shown in light blue, the "β-flap" motif is demonstrated in pink. (C) The crystal structure of a fragment of the CROPs domain of TcdB (TcdB1834-2099; PDB-ID 4NP4, Orth *et al.*, 2014) is presented in light green. Two repeats are highlighted in dark green and orange, respectively.

6.3. Comparison of the homologous protein toxins TcdA and TcdB

The overall sequence identity of TcdA and TcdB is around 49%. Both toxins share the same structural organization and can be separated into four individual domains (Jank & Aktories, 2008). A sequence and structural alignment of the individual TcdA and TcdB domains reveals high similarity of three domains of toxin A and B.

So far, no crystal structures of full-length TcdA and TcdB have been published. Only crystal structures of the glucosyltransferase domain, the autoprotease domain and fragments of the CROPs domain are solved for both toxins. For TcdA, a crystal structure of TcdA1-1838 including the intermediate translocation domain has been published by another group during the progress of this thesis. This crystal structure is not available for toxin B. Nevertheless, a Phyre model of the TcdB translocation domain was generated in this thesis based on homology of the TcdB domain to the TcdA translocation domain. The structural similarity could be proved by SAXS analysis (chapter 6.5.3). The individual crystal structures as well as the Phyre model of the TcdB translocation domain and the crystal structure of the TcdA translocation domain will be analyzed in this chapter regarding their sequence and structural similarity.

Table 16: Sequence and structural similarity of the individual domains of TcdA and TcdB. The sequence similarity was calculated using EMBOSS Needle (Needleman & Wunsch, 1970), the structural similarity was determined by a structural alignment performed in Coot (Emsley *et al*., 2010).

Domain	Number aa	Sequence identity of TcdA and TcdB domains	Structural similarity (r.m.s.d.) of TcdA and TcdB domains
GTD	TcdA-GTD 555 aa (PDB 3SRZ) TcdB-GTD 543 aa (PDB 2BVL)	50.6%	1.10 Å (529 aligned residues)
APD	TcdA-APD 267 aa (PDB 3HO6) TcdB-APD 254 aa (PDB 3PEE)	55.2%	1.00 Å (239 aligned residues)
IMD	TcdA-IMD 1021 aa (PDB 4R04) TcdB-IMD 1033 aa (Phyre model)	46.3%	6.01 Å (699 aligned residues)
CROPs	TcdA-CROPs 878 aa (Phyre model) TcdB-CROPs 533 aa (Phyre model)	28.0%	2.32 Å (143 aligned residues)

The TcdA and TcdB glucosyltransferase domains (PDB-ID 3SRZ, Pruitt *et al*., 2012; PDB-ID 2BVL, Reinert *et al*., 2005) share a sequence identity of 50.6% and the structures could be aligned with an r.m.s.d. of 1.10 Å. This indicates that the fold of both domains is almost identical with only minor differences in their sequence. A structure-based sequence alignment (shown in Fig. 81 in the supplement) shows that these differences are mainly located in loop regions and that the secondary structure elements as well as the active site composition of the TcdA and TcdB glucosyltransferase domain are highly conserved.

Beside the glucosyltransferase domains, also the autoprotease domains of TcdA and TcdB (PDB-ID 3HO6, Pruitt *et al*., 2006; PDB-ID 3PEE, Shen *et al*., 2011) show a high sequence

and structural similarity. The sequence identity of both domains is 55.2% and both structures can be aligned with an r.m.s.d. of 1.00 Å, indicating an almost identical protein fold. A structure-based sequence alignment is shown in Fig. 82 in the supplement. Patches of conserved residues are distributed over the whole protein structure and particularly all residues surrounding the active site, the $InsP_6$ binding site and the β-flap are highly conserved.

The only domain of TcdB that has structurally not been characterized so far is the intermediate translocation domain. The intermediate domain of TcdA and TcdB share a sequence identity of 46%. Due to their high sequence identity, it can be assumed that both domains share a similar protein fold. Therefore, a structural model of the TcdB translocation domain was calculated using *Phyre2* (Kelley *et al.*, 2015) based on the TcdA translocation domain as homology model. The crystal structure of the TcdA translocation domain (PDB-Code 4R04; Chumbler *et al.*, 2016) and the Phyre model of the TcdB translocation domain can be aligned with an r.m.s.d. of 6.01 Å (Fig. 83 in the supplement). The central part of both domains can be aligned without striking deviations, but structural discrepancies between both domains are visible in the terminal regions of the domains. As the TcdA translocation domain was crystallized in presence of the neighboring autoprotease and glucosyltransferase domain, its structure could show slight differences in the orientation of its termini in contrast to a crystal structure without adjacent domains. The TcdB translocation domain termini could hence also undergo conformational changes in the presence of the neighboring domains, which might explain the high r.m.s.d. value upon alignment with the TcdA translocation domain. Another indicator for the high r.m.s.d. is the fact that a Phyre model of the TcdB translocation domain was used for the structural alignment. A crystal structure of this domain would be necessary to reveal the exact molecular structure and orientation of this domain and could result in a lower r.m.s.d. value. The structure-based sequence alignment of the TcdA and TcdB intermediate domain is shown in Fig. 14.

In contrast to the above-mentioned highly similar domains of TcdA and TcdB, the C-terminal CROPs domain shows striking differences between toxin A and B (PDB-ID 2QJ6, Albesa-Jové *et al.*, 2011; PDB-ID 4NP4, Orth *et al.*, 2014). Their CROPs domains vary in their length, as well as in their amino acid composition. The TcdA CROPs domain is more than 300 residues longer than the TcdB CROPs domain and both domains share a sequence identity of only 28%. The structural and especially the sequence differences between both homologous protein toxins could explain the fact that both toxins bind to different receptors on the host cell surface (Na *et al.*, 2008; LaFrance *et al.*, 2015). Low-resolution EM-studies (Pruitt *et al.*, 2010) revealed that the overall shape and orientation of the TcdA CROPs domain is a long elongated tail protruding from the rest of the toxin structure. In contrast to this, the density for the TcdB CROPs domain was difficult to interpret, indicating that the TcdB CROPs domain is highly

flexible. This has also been shown by SAXS experiments of a TcdB fragment including the CROPs domain in this thesis (chapter 6.5.3.2).

Overall, it can be summarized that full-length toxin A and toxin B share a similar structure with exception of their C-terminal CROPs domain. A full-length crystal structure or a SAXS envelope including all domains of both toxins would be essential to prove this hypothesis and to analyze the structural similarity of the homologous toxins of *Clostridium difficile* in atomic detail.

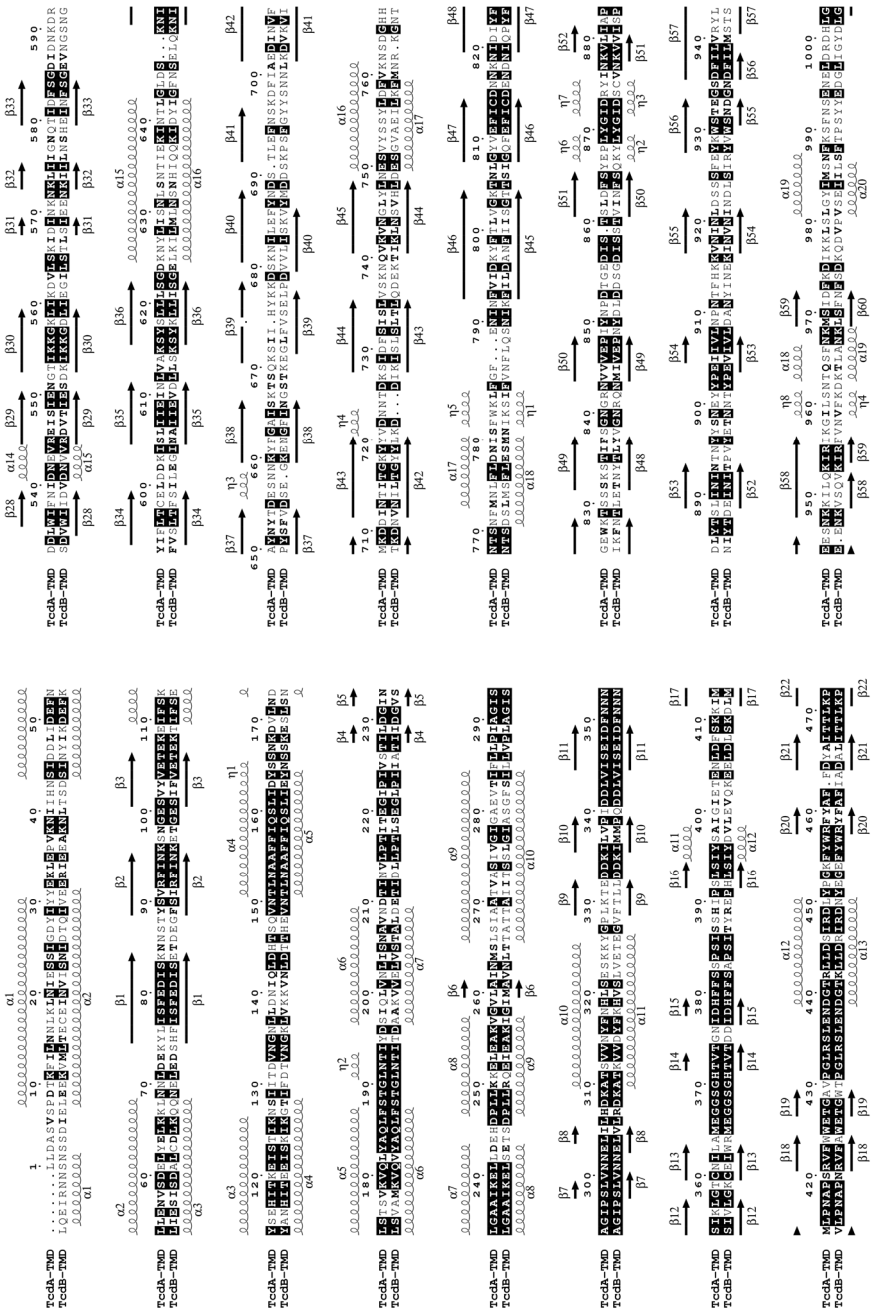

Figure 14: Structure-based sequence alignment of the TcdA and TcdB intermediate domain. The sequence alignment was calculated by PromalS3D and edited by Espript. The aligned protein sequences of TcdA and TcdB are shown with conserved residues highlighted by a black box and similar amino acids are highlighted in bold. The secondary structure of the TcdA intermediate domain (PDB-ID 4R04; Chumbler *et al.*, 2016) is depicted above the alignment, the secondary structure of a Phyre model of the TcdB intermediate domain is illustrated under the sequence alignment. α-helices are illustrated by spirals and β-sheets as arrows. The secondary structure elements are numbered consecutively.

6.4. Aim of this project

At the beginning of this thesis, crystal structures of the intermediate domain of TcdB and full-length TcdB were not available. For a better understanding of translocation and pore formation of Toxin B from *Clostridium difficile*, an atomic structure of its pore-forming domain is essential. The aim of this part of the thesis was therefore the structural analysis of the TcdB intermediate domain by X-ray crystallography and small angle X-ray scattering (SAXS).

6.5. Structural analysis of TcdB from *Clostridium difficile*

6.5.1. Establishment of a purification protocol for TcdB fragments

Several expression constructs of the intermediate domain of TcdB were cloned from genomic *Clostridium difficile* CD630 DNA, as well as from synthetic DNA Codon-optimized for expression in *E. coli*. All TcdB constructs are summarized in table 35 - table 37 in the supplement. Test expression of all constructs was performed analogous to chapter 5.5, but expression of TcdB constructs in *E. coli* was problematic due to solubility problems and degradation of the recombinant proteins. To overcome degradation problems, expression of some constructs was performed in *Bacillus megaterium*. This bacterium is similar to *Clostridium difficile* and does not possess alkaline proteases, which could degrade recombinant gene products (Vary, 1994). It is therefore a suitable host for expression of *Clostridium difficil*e toxins and toxin fragments (Yang *et al*., 2008; Burger *et al*., 2003). The construct that yielded the highest amount of soluble and stable TcdB was pHIS1522-TcdB543-2366. This construct includes the autoprotease domain, the intermediate domain, as well as the CROPs domain of TcdB. It is fused to a non-cleavable N-terminal Strep- and a C-terminal His_6-tag. Due to the intrinsic autoproteolytic activity of TcdB, the N-terminal Strep-tag was cleaved off during expression and purification of the construct via Strep-affinity chromatography was hence not possible. Therefore, the catalytically active residue Cys698 was inactivated by mutation to an alanine residue using site-directed mutagenesis to prevent autocatalytic cleavage of the N-terminal Strep-tag.

The resulting modified construct Strep-TcdB543-2366(C698A)-His_6 was expressed in *Bacillus megaterium* at 37 °C in LB-medium, the detailed expression protocol is elucidated in chapter 5.6.1. After cell harvesting, the cell pellet was resuspended in buffer (20 mM Tris/HCl pH 8.0, 300 mM NaCl) supplemented with DNAse and protease inhibitor. The first purification step involved a Ni(II)-IMAC (binding buffer 20 mM Tris/HCl pH 8.0, 300 mM NaCl and elution buffer including 500 mM imidazole). The construct was eluted from the His-Trap column using a linear gradient from 0-500 mM imidazole. The fractions containing the TcdB construct were dialyzed against 20 mM Tris/HCl pH 8.0, 300 mM NaCl at 4 °C overnight to get rid of remaining

imidazole. Purification of the TcdB fragment involved a second affinity chromatography step using a Strep-Tactin column (20 mM Tris/HCl pH 8.0, 300 mM NaCl as binding buffer and 20 mM Tris/HCl pH 8.0, 300 mM NaCl, 5 mM Desthiobiotin as elution buffer). A final size exclusion chromatography step was performed to remove potential aggregates using a S200 16/60 column in 20 mM Tris/HCl pH 8.0, 300 mM NaCl. The final protein yield was approximately 0.2 mg protein per liter cell culture. A schematic overview of the purification protocol, as well as the chromatogram of the final size exclusion chromatography step and an SDS-PAGE with the respective protein fractions are shown in Fig. 15. A further stable TcdB fragment (TcdB800-1834) could be purified in this thesis that was used for SAXS experiments (chapter 6.5.3.1).

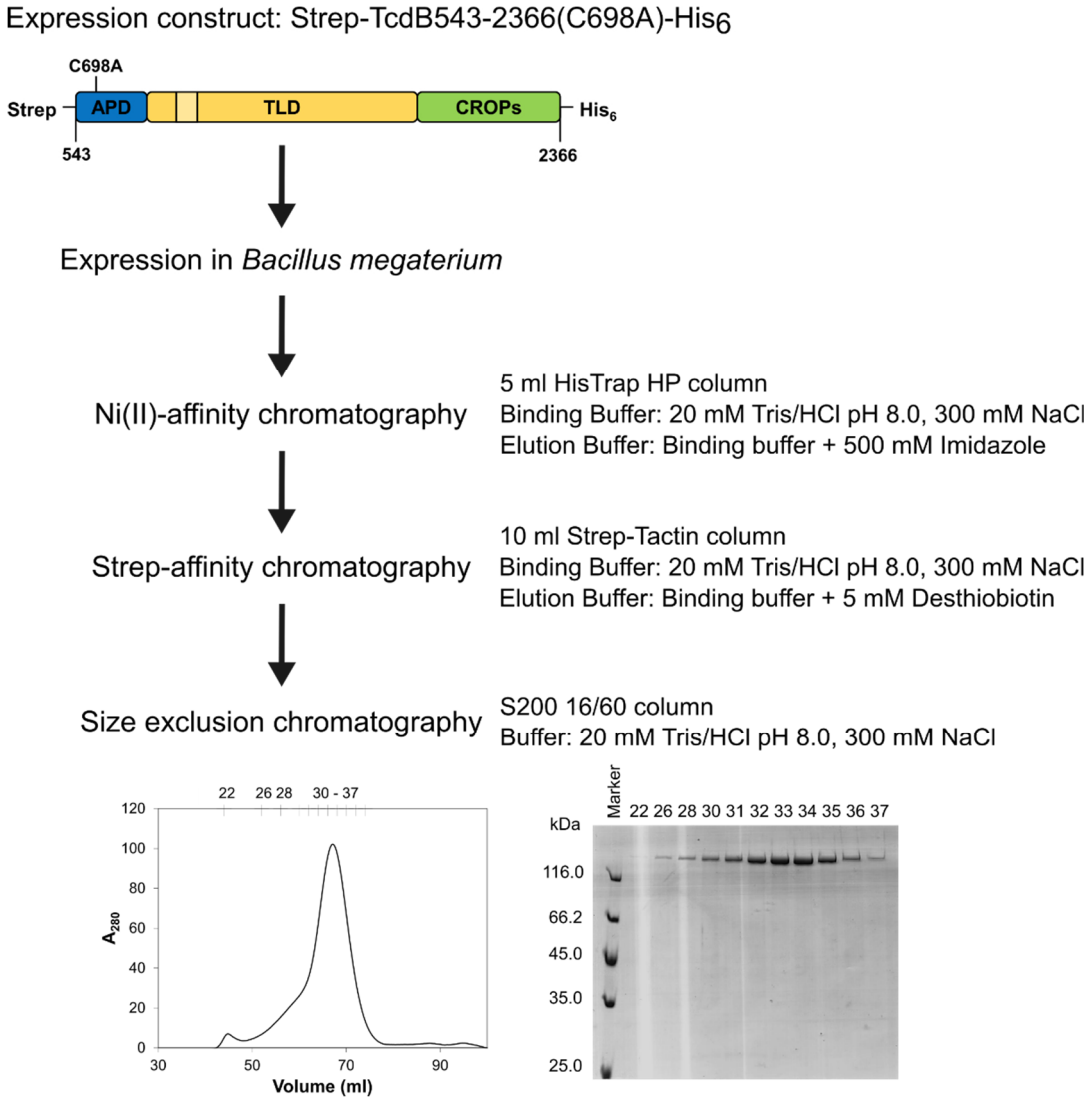

Figure 15: Production and purification of Strep-TcdB543-2366(C698A)-His_6. The chromatogram and the corresponding SDS-PAGE of the final size exclusion chromatography are depicted in the bottom part of the scheme. A protein standard is shown in the first lane of the SDS-PAGE for comparison of the molecular weight of the purified construct (theoretical molecular weight ~ 210 kDa).

6.5.2. Crystallographic and EM-experiments of TcdB fragments

TcdB fragments that showed sufficient purity after protein purification were used for crystallization experiments. Several commercially available screens in combination with different protein concentrations were used. Unfortunately, no crystal growth could be obtained for any TcdB construct during this thesis work.

Besides protein crystallization, the construct Strep-TcdB543-2366(C698A)-His_6 was also used for cryo-EM experiments. Initial negative stain screenings using uranyl acetate were performed at Prof. Dr. Thomas Marlovits' lab (IMBA, Wien) to analyze whether the quality of the protein sample is suitable for following cryo-EM experiments. As the initial screenings revealed that many aggregates are present in the protein sample (Fig. 16), an improvement of sample quality should be performed prior to further cryo-EM experiments.

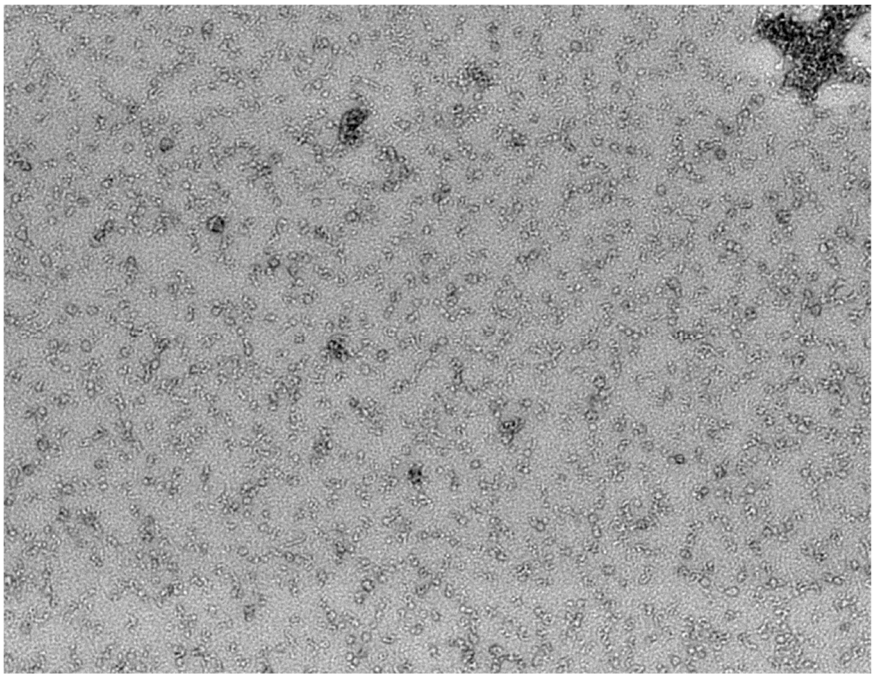

Figure 16: Initial negative staining experiments of Strep-TcdB543-2366(C698A)-His_6 with uranyl acetate (1:20 dilution) reveal the presence of aggregates in the protein sample.

6.5.3. SAXS analysis of TcdB fragments

As both protein crystallization and initial electron microscopy experiments failed to reveal structural information about TcdB, SAXS experiments were performed to analyze the overall shape of TcdB and its oligomeric state in solution. SAXS data for two constructs (Strep-TcdB543-2366-His_6(C698A) and TcdB800-1834) were collected at BM29 ESRF (Grenoble, France; Pernot *et al.*, 2013). Collection statistics are summarized in table 17.

Table 17: Summary of SAXS measurement parameters for TcdB constructs

Construct	Strep-TcdB543-2366 (C698A)-His_6	TcdB800-1834
Buffer	20 mM Tris/HCl pH 8.0, 300 mM NaCl	20 mM Tris/HCl pH 8.0, 300 mM NaCl, 10% glycerol
Protein concentration	0.57 mg/ml	0.5 mg/ml
Number of experimental data points	1040	1040
Dro	0.000	0.075
R_a	1.800	1.400
R_g	66.09 nm	43.71 nm
Chi value	346.80	2.12
Volume	287790 nm^3	157475 nm^3

6.5.3.1. SAXS envelope of TcdB800-1834

The construct TcdB800-1834 is composed of the intermediate translocation domain of TcdB. Solubility and affinity tags, which were attached to this construct for expression and purification, were cleaved off prior to SAXS experiments. The SAXS experiments of TcdB800-1834 were performed to get insights into the overall shape, conformation and orientation of the translocation domain of TcdB, as no crystal structure of this domain was available at the onset of this thesis.

The fit of the experimental SAXS data and a theoretically calculated scattering curve for monomeric TcdB800-1834 is shown in Fig. 17A. A Phyre model of the TcdB translocation domain was generated based on homology to the TcdA translocation domain and was used for fitting the experimental data. A *X*-value of 2.12 indicates a good fit of experimental and theoretical data for monomeric TcdB800-1834 and indicates that the toxin fragment TcdB800-1834 is monomeric in solution. The Phyre model of monomeric TcdB800-1834 was afterwards fitted into the calculated SAXS envelope (Fig. 17B).

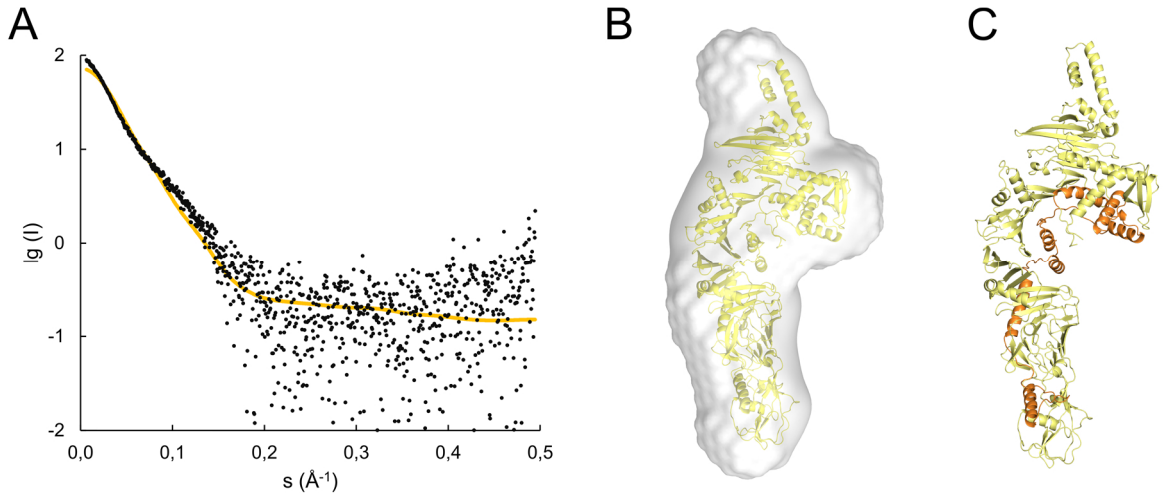

Figure 17: SAXS analysis of TcdB800-1834. (A) Fit of the experimental data (black dots) and a theoretical SAXS curve for monomeric TcdB800-1834 (orange curve). (B) Fit of a Phyre model of TcdB800-1834 (yellow cartoon representation) and the calculated SAXS envelope for monomeric TcdB800-1834. (C) Mapping of the hydrophobic region (orange), which is probably involved in pore formation, on the TcdB intermediate domain (yellow).

The Phyre model of the TcdB translocation domain (TcdB800-1834) shows no homology to other known protein structures besides TcdA from *Clostridium difficile*. It consists mainly of β-sheets. The hydrophobic region of TcdB (TcdB980-1107) (Zhang *et al.*, 2014) that was shown to be probably involved in pore formation of the toxin was mapped on the Phyre model of the TcdB translocation domain (Fig. 17C). It shows that this hydrophobic region consists mainly of α-helical parts that wrap around the central β-sheets and reach from one end of the structure to the other. A long loop is located at the terminal part of this hydrophobic region, which is followed by a β-hairpin that contains the identified essential residues required for pore formation. It is assumed that these residues insert into the endosomal membrane (Zhang *et al.*, 2014). Due to their location in the flexible loop and hairpin, their insertion into the endosomal membrane could be easily achieved by small conformational changes in the translocation domain.

Because of the good fit of the TcdB Phyre model into the calculated SAXS envelope, it can be assumed that the theoretical Phyre model of the TcdB translocation domain reveals the conformation of the TcdB translocation domain in solution and is highly similar to the TcdA intermediate domain. However, only a crystal structure of the TcdB intermediate domain will reveal its exact molecular structure and conformation.

6.5.3.2. SAXS envelope of Strep-TcdB543-2366(C698A)-His_6

The orientation and shape of the TcdB translocation domain could be revealed by the SAXS envelope of TcdB800-1834 (chapter 6.5.3.1). To analyze its relative position with respect to the neighboring domains, additional SAXS experiments of the construct Strep-TcdB543-2366(C698A)-His_6 have been performed. This fragment includes the autoprotease domain, the translocation and the CROPs domain of TcdB. The affinity tags used for purification could not be removed prior to the SAXS experiments, but due to their small size they should not hamper the measurements.

Prior to the SAXS measurements, a theoretical model of TcdB was prepared using *Phyre2* (Kelley *et al.*, 2015) based on homology to the TcdA crystal structure (PDB-ID 4R04; Chumbler *et al.*, 2016). This model was also used for a fit into the calculated SAXS envelope of Strep-TcdB543-2366(C698A)-His_6 (Fig. 18B). The fit of experimental SAXS data and the theoretically calculated scattering curve for the monomeric TcdB model is shown in Fig. 18A. A high *X*-value of 28.1 indicates a bad fit of experimental data and the theoretical model of monomeric TcdB. The individual TcdB domains were afterwards fitted separately into the calculated SAXS envelope (Fig. 18B). The crystal structure of the TcdB autoprotease domain (PDB-ID 3PEE; Shen *et al.*, 2011) is shown in blue. For the fit of the intermediate translocation domain (TcdB800-1834) the calculated Phyre model (colored in yellow) was used in the same orientation that was revealed by the SAXS envelope of TcdB800-1834 (chapter 6.5.3.1). Due to the lack of a crystal structure of the full-length TcdB CROPs domain the structure of a CROPs fragments (PDB-ID 4NP4; Orth *et al.*, 2004) was fitted two times into the SAXS envelope to model the full-length CROPs domain. Because this is only a manually modelled representation of the TcdB fragment, this fit of the individual TcdB domains into the SAXS envelope has to be interpreted with care.

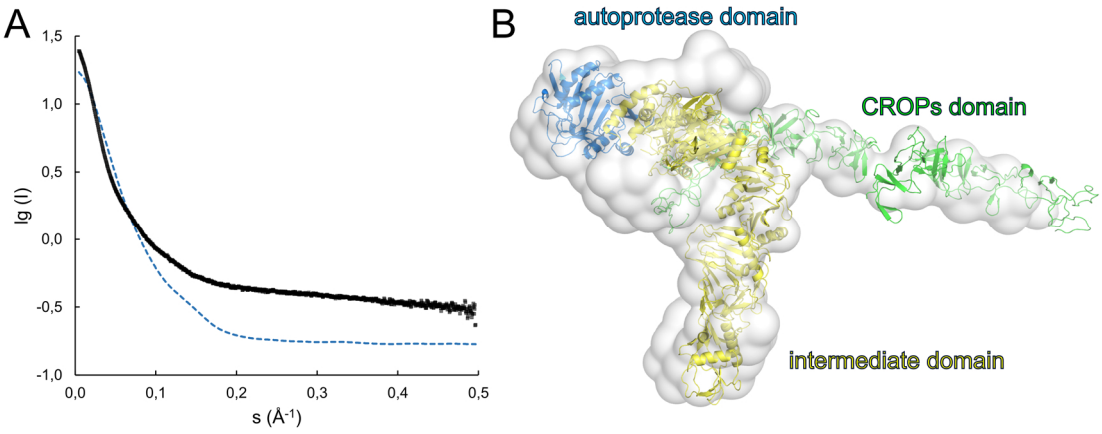

Figure 18: SAXS analysis of Strep-TcdB543-2366(C698A)-His_6. (A) Fit of experimental data (black dots) and a theoretically calculated scattering curve for monomeric Strep-TcdB543-2366(C698A)-His_6 (dashed blue line). (B) Fit of the individual TcdB domains into the calculated SAXS envelope for monomeric Strep-TcdB543-2366(C698A)-His_6. The domains are shown in cartoon representation with the autoprotease domain in blue, the Phyre model of the intermediate domain in yellow and the CROPs domain in green.

The TcdA crystal structure (PDB-ID 4R04; Chumbler *et al*., 2016), which was published by another group during the progress of this thesis, revealed that the TcdA translocation domain starts at the C-terminal end of the APD domain and folds in a horseshoe-shape back to the APD domain. Due to high sequence similarity of TcdA and TcdB, it is assumed that the same is true for the TcdB structure. Therefore, it is highly possible that the TcdB CROPs domain starts between translocation and APD domain at the base of the APD domain. Also the EM envelope of TcdB (Pruitt *et al*., 2010) shows that the CROPs domain of TcdB is probably located in near proximity of the APD.

The SAXS envelope of Strep-TcdB543-2366(C698A)-His_6 indicates a high flexibility of the CROPs domain, as SAXS density for this domain is incomplete and only poorly resolved. It is usually difficult to calculate a SAXS envelope for highly flexible proteins or proteins that can be present in different conformational states (Putnam *et al*., 2007). It has already been shown by EM experiments (Pruitt *et al*., 2010) that the CROPs domain of TcdB can be present in different conformational states, while the rest of the protein does not change its orientation. Due to its high flexibility and the fact that the CROPs domain was modelled manually into the calculated SAXS envelope, it can be assumed that the orientation of the CROPs domain as illustrated in Fig. 18B will most probably not depict its actual conformation in solution. A crystal structure of the full-length CROPs domain or full-length TcdB would be helpful to discuss the orientation of the individual TcdB domains respective to each other. To summarize, the SAXS envelopes of TcdB fragments including the autoprotease, the intermediate and the CROPs domain of TcdB, which were calculated in this thesis, revealed that the TcdB intermediate translocation domain is highly similar to the TcdA domain regarding the overall shape of the domain.

7. Structural analysis of essential proteins

7.1. Introduction

7.1.1. The importance of structure determination of uncharacterized proteins

The Protein Data Bank (PDB; Berman *et al*., 2000) is a database for crystal structures of biological macromolecules. Currently, 115.639 protein structures determined by X-ray crystallography are deposited in the PDB, which include only 120 protein structures of *Clostridium difficile* (status 22.02.18). As the genome of *Clostridium difficile* 630 encodes over 3.700 proteins (Sebaihia *et al*., 2006), the amount of published structural information on these proteins is limited. Due to the inevitable need for new therapeutic treatment methods for CDI (Smits *et al*., 2016), the structural analysis of uncharacterized *C. difficile* proteins is an unavoidable requirement. Discovering new drug targets is only feasible, when the metabolic pathways and proteins involved in central processes inside the pathogen are well characterized (Lindsay, 2005). Structural and functional analysis of essential proteins of *C. difficile* has to go hand in hand to reveal insights into the bacterium and to uncover new potential drug targets.

The aim of this project hence involved the structural and functional analysis of proteins from *Clostridium difficile* that are involved in central cellular processes. A detailed understanding of these proteins and the underlying mechanisms will guide the way to the development of new treatment methods for *Clostridium difficile* associated infections. Determination of novel protein structures is not only essential for a detailed atomic characterization but can also aid at functional characterization (Watson *et al*., 2005). A closer look at highly resolved crystal structures might identify potential binding cavities for protein ligands. By analysis of the surrounding amino acids and comparison to structurally related proteins, one could thus derive hypotheses about the potential molecular function (Wild *et al*., 2004), as conserved active site residues often indicate similar protein functions. Based on such findings, future biochemical experiments can be designed and initial *in silico* experiments, such as docking (Hermann *et al*., 2006), can be performed (Zhang & Kim, 2003). Nevertheless, it has to be considered that some proteins undergo conformational changes upon ligand binding, which makes docking experiments into apo-protein structures difficult and hence requires additional experiments for function determination (Hermann *et al*., 2007).

The number of protein folds present in nature is limited (Chothia, 1976) and the amount of protein crystal structures available in the PDB is already extremely high. Therefore, it is likely that one finds a homologous protein with the same fold after solving the crystal structure of an uncharacterized protein. A comparison of the new structure with homologous functionally and structurally characterized proteins can hence give information on their potential molecular

function. On the other hand, analysis of structurally uncharacterized proteins is essential to detect novel protein folds that have not been deposited in the PDB so far (Wild *et al.*, 2004).
Besides a detailed structural characterization, analysis of the genomic environment of an uncharacterized protein can also reveal information about the function of the encoded protein, as proteins that are involved in the same metabolic pathway are often also located in the same genetic operon (Zhao *et al.*, 2013).
As structurally similar proteins might have different molecular functions in some cases (Finkelstein *et al.*, 1993), additional experiments always have to be performed in order to prove a hypothetical function that was derived from the protein structure. Due to the difficulty of assigning a function to novel protein structures, many protein structures with unknown function are deposited in the PDB.

7.1.2. Selection of uncharacterized essential proteins of *Clostridium difficile*

The selection of uncharacterized proteins for this project was based on the work of Dembek *et al.*, 2015. They identified several essential genes of the highly virulent epidemic *Clostridium difficile* strain R20291 using transposon-directed insertion site sequencing (TraDIS). Among these are 798 genes that are involved in spore production and 404 genes that are essential for growth of *Clostridium difficile in vitro*. Some of the encoded proteins could be functionally annotated due to homology to characterized proteins from related bacterial strains. Besides that, many proteins without annotated function have been identified that could serve as potential targets for this thesis. The distribution of characterized and uncharacterized proteins among all identified essential proteins is illustrated in Fig. 19. Only 242 of the 404 genes essential for growth have an annotated function. 109 genes have “putative” functions, 33 are annotated as hypothetical proteins and 20 genes encode uncharacterized proteins. Among the 798 genes that influence spore production are 349 with annotated function, 334 with “putative” function, 96 hypothetical proteins and 19 uncharacterized proteins. The uncharacterized proteins are of special interest for this project in regards of finding new drug targets and novel protein folds.

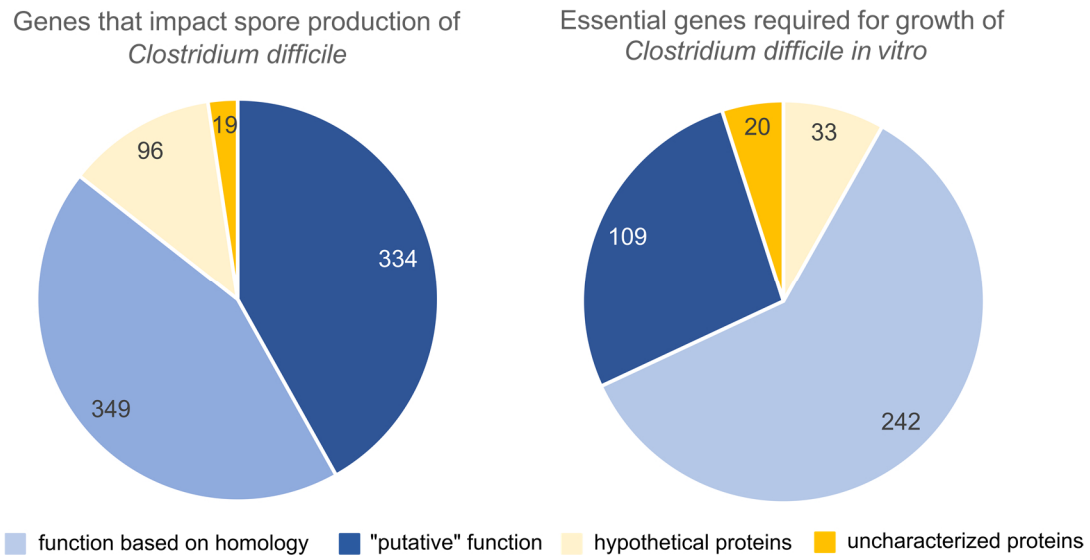

Figure 19: Distribution of the genes that were identified by Dembek *et al.*, 2015 to be essential for growth or sporulation of *Clostridium difficile* R20291 based on their annotated function.

To select proteins for this part of the thesis, a BLAST search was performed first to find the respective homologues of the essential genes of *C. difficile* R20291 in the laboratory strain *C. difficile* 630. Several cut-off criteria were applied to discard proteins that are less well suited for structural characterization by X-ray crystallography, e.g. because of their content of transmembrane helices, their molecular weight or structural predictability etc. Proteins that were selected here contained 200 – 400 amino acids and no transmembrane helices, as these would hamper protein crystallization. Furthermore, their structure should not be predictable by common structure prediction servers. To search for structurally uncharacterized and unpredictable proteins, the *PhyreBug* server (unpublished) was developed by Dr. Joachim Reichelt. Using this server, whole genomes can be scanned combined with information about gene length, content of transmembrane helices and information about operon structure. Moreover, the structure prediction of the encoded protein by *Phyre2* is provided. It was hence possible to discard proteins with known structure and high homology to structurally characterized proteins directly.

The proteins that were finally selected based on the above-mentioned criteria are summarized in table 18. The genes encoding these proteins were synthetically codon-optimized for expression in *E. coli* and purchased from Life Technologies GmbH (Darmstadt, Germany). They were cloned in various expression vectors by SLIC-cloning as well as traditional cloning using restriction enzymes. Test expression in *E. coli* was performed with all constructs and the best condition was used for large-scale expression. Proteins were purified using affinity chromatography, ion exchange and size-exclusion chromatography to yield pure protein that was subsequently used for protein crystallization. Structural analysis was performed by X-ray crystallography and small angle X-ray scattering. The determined protein structures were then compared to published structures in the PDB to search for structurally related proteins and to

derive ideas about their potential function. A schematic overview of the general workflow together with achieved results for each protein is summarized in Fig. 54.

Table 18: Summary of essential proteins of *Clostridium difficile* that were selected for this project.

Protein	Molecular Weight (kDa)	Number of amino acids	Theoretical pI	Involved in
CD1067	44.76	405	4.78	sporulation
CD1219	39.51	355	5.96	growth
CD1823	29.19	247	9.10	sporulation
CD2589	34.26	293	4.89	growth
CD2752	25.63	221	5.21	sporulation

7.2. The uncharacterized protein CD1219 of *Clostridium difficile*

CD630_12190 (CD1219) of *Clostridium difficile* is a 39.5 kDa protein with unknown function and structure. In 2015, Dembek *et al*. showed that its homologous gene CDR20291_1057 from *C. difficile* strain R20291 is essential for growth of *C. difficile in vitro*. Thus, it can be assumed that CD1219 might also be involved in central cellular processes and determination of its crystal structure would facilitate a detailed understanding of its structure and molecular function.

7.2.1. Cloning, expression and purification of CD1219

Bioinformatic analysis and secondary structure prediction of CD1219 by *PSIPRED 3.3* (McGuffin *et al*., 2000) revealed that the protein does not contain disordered regions at its N- and C-terminus, respectively. Therefore, full-length CD1219 was used for cloning experiments. To overcome potential expression problems, the DNA encoding CD1219 was codon-optimized for expression in *E. coli* (synthetic DNA ordered at LifeTechnologies, Darmstadt, Germany). Synthetic CD1219 DNA was cloned in a pET-Sumo vector by SLIC cloning, resulting in CD1219 fused to an N-terminal His_6-Sumo tag followed by a cleavage site for Sumo-protease. After test expression in several *E. coli* strains, the fusion protein was solubly expressed in *E. coli* Rosetta2(DE3) cells in TB-medium after induction with 1 mM IPTG at 20 °C overnight. Native CD1219 was purified via 4-step purification (table 12). After a Ni(II)-IMAC, the His_6-Sumo tag was cleaved by Sumo-protease overnight. To get rid of cleaved tag and protease, a second Ni(II)-IMAC was performed, followed by anion exchange chromatography and a size exclusion chromatography. The final yield of pure protein after size exclusion chromatography was 11 mg from 1 L cell culture. The chromatogram of the size exclusion chromatography of native CD1219 and the corresponding SDS-PAGE are shown in Fig. 20. The protein elutes as a dimer and the peak shows a small shoulder indicating that monomeric

CD1219 is still present in the sample. These fractions of monomeric CD1219 were hence discarded prior to further experiments.

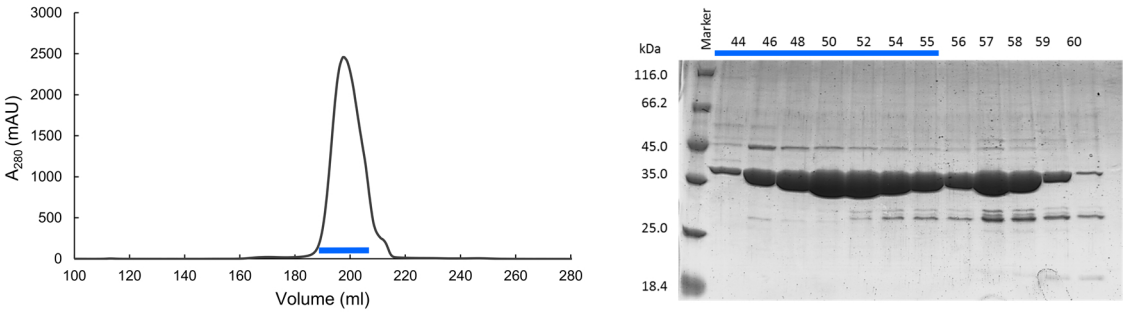

Figure 20: Final size exclusion chromatography of native CD1219. The elution profile of the final size exclusion chromatography of native CD1219 using a S200 26/60 column is shown in the left panel. The corresponding SDS-PAGE with eluted fractions is shown in the right figure. The blue bar represents the fractions that were pooled after size exclusion chromatography and were used for crystallization and SAXS experiments. A protein standard is shown in the left lane of the SDS-PAGE for an estimation of the molecular weight of the purified protein. The theoretical molecular weight of CD1219 is 39.5 kDa.

7.2.2. Crystallization and structure determination of CD1219

Initial crystallization screens (Core I-IV, Morpheus, Index, ProteinComplex) were set up with native CD1219 in different concentrations (20 mg/ml, 10 mg/ml, 5 mg/ml) immediately after purification. Crystal growth was obtained in various conditions already after 12 hours. Datasets of native CD1219 crystals were collected at PXII at SLS (Villigen, Switzerland; Fuchs *et al.*, 2014). Due to the lack of structural homologs of CD1219, molecular replacement was not possible for structure determination. Soaking heavy atoms into the native crystals was not successful due to crystal cracking. Therefore, SeMet-labeled CD1219 was produced in *E. coli* Rosetta2(DE3) analogous to the protocol in chapter 5.6.2.2. Purification of SeMet-labeled CD1219 was performed analogously to native CD1219. The final size exclusion chromatography elution profile of SeMet-labeled CD1219 and the corresponding SDS-PAGE is shown in Fig. 21. Crystallization screens were set up analogously to native CD12129 and crystals of SeMet-CD1219 appeared after 12 hours in various conditions. High-redundancy datasets of SeMet-CD1219 crystals were collected at PXII, SLS. Exemplary CD1219 crystals are shown in Fig. 22. Data collection statistics of native CD1219 and SeMet-CD1219 crystals are summarized in table 19.

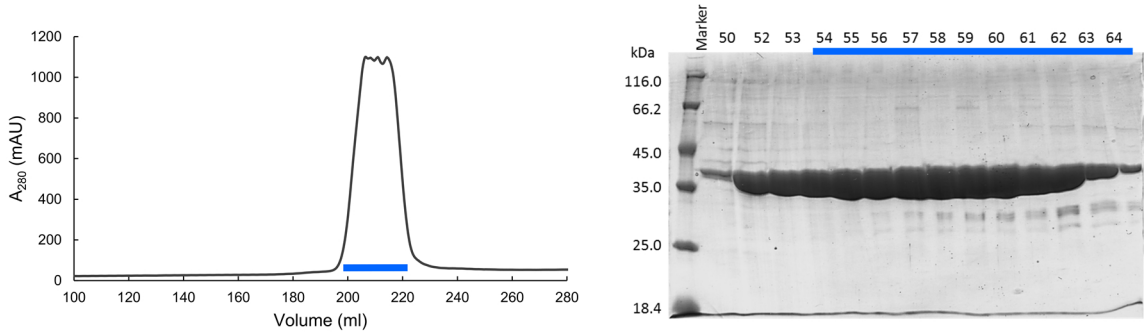

Figure 21: Final size exclusion chromatography of SeMet-labeled CD1219. The elution profile of the final size exclusion chromatography (using a S200 26/60) is shown in the left figure. The corresponding SDS-PAGE with eluted protein fractions is shown in the right figure. A protein standard was loaded in the left lane of the SDS-PAGE to estimate the molecular weight of the purified protein. The theoretical molecular weight of SeMet-labeled CD1219 is approximately 40 kDa. The blue bar represents the fractions that were pooled after size exclusion chromatography and were used for crystallization setups of SeMet-labeled CD1219.

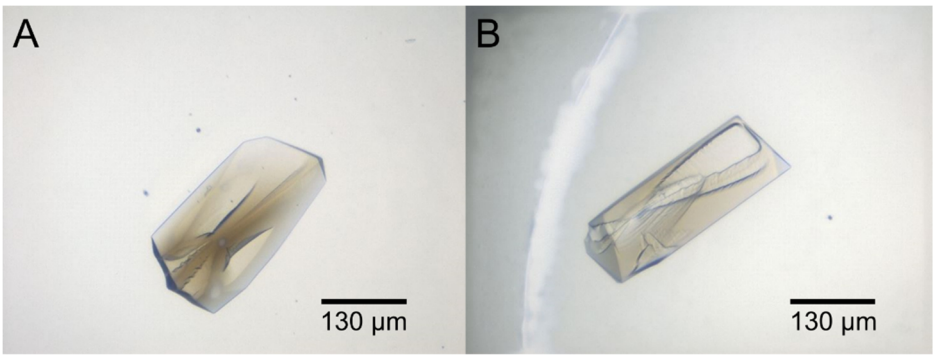

Figure 22: Exemplary crystals of native CD1219 (A) and SeMet-labeled CD1219 (B). (A) Crystal of native CD1219 in Morpheus Screen appeared 2 days after setup in the crystallization condition 5 mg/ml CD1219, 30% P550MME-P20K, 0.1 M Morpheus buffer 1 pH 6.5, 0.09 M Morpheus Nitrate-Phosphate-Sulfate. (B) Crystal of SeMet-labeled CD1219 in Morpheus Screen emerged 2 days after setup in the crystallization condition 5 mg/ml SeMet-CD1219, 30% P550MME-P20K, 0.1 M Morpheus buffer 3 pH 8.5, 0.09 M Morpheus Nitrate-Phosphate-Sulfate.

Table 19: Data collection statistics for a native CD1219 dataset and a SeMet dataset.

Dataset	**CD1219_native**	**CD1219_SeMet**
Wavelength (Å) / beamline	1.000 / SLS, PXII	0.978 / SLS, PXII
Resolution range (Å)	48.29-1.80 (1.84-1.80)	48.33-1.80 (1.84-1.80)
Space group	$P2_12_12_1$	$P3_121$
Unit cell parameters (Å)	56.1 102.8 141.2	104.06 104.06 130.56
(°)	90 90 90	90 90 120
Mosaicity (°)	0.20	0.05
Total No. of measured reflections	1019887 (52958)	3080270 (171377)
Unique reflections	76500 (4464)	76151 (4471)
Multiplicity	13.3 (11.9)	40.4 (38.3)
Mean I/σ(I)	18.6 (1.4)	31.7 (3.2)
Completeness (%)	100 (100)	100 (100)
R_{meas} (%)	11.0 (202.3)	9.9 (156.8)
$R_{p.i.m.}$ (%)	4.1 (81.4)	2.2 (35.3)
CC1/2 (%)	99.9 (53.2)	100 (85.3)

Native CD1219 crystallized in spacegroup $P2_12_12_1$ with two molecules in the asymmetric unit (AU). Native crystals diffracted to 1.8 Å. SeMet-labeled CD1219 crystallized in spacegroup $P3_121$ with two molecules in the AU. Crystals of SeMet-CD1219 also diffracted to 1.8 Å. The crystal structure of CD1219 was solved by single-wavelength anomalous diffraction (SAD) using one high-redundant SeMet dataset and *Autosol* for structure determination and data processing. An initial model of CD1219 was prepared using *Autobuild*. Further manual model building was performed with *Coot* and *Phenix.refine* was used for structure refinement. As native and anomalous CD1219 crystallized in different spacegroups, the final CD1219 model was used for molecular replacement using the native dataset of CD1219 by *Phenix.phaser*, but no differences between the structure of native and SeMet-CD1219 were visible. For the rest of this thesis, the native CD1219 structure was used for structural analysis. The refinement statistics for the structure of native CD1219 are shown in table 20 and the quality of electron density is exemplarily depicted in Fig. 23.

Table 20: Refinement statistics of CD1219.

Dataset	**CD1219_native (Refine31)**
Resolution Range (Å)	48.29 - 1.8 (1.864 - 1.8)
R_{work} (%)	16.0 (26.0)
R_{free} (%)	18.9 (30.2)
No. of non-H atoms	6172
Protein	5547
Ligand	62
Water	563
R.m.s. deviation	
Bonds (Å)	0.008
Angles (°)	0.87
Average B factors (Å^2)	
Protein	39
Ligand	70
Water	48
Ramachandran plot	
Favored regions (%)	97.29
Outliers (%)	0.00
MolProbity score	1.35

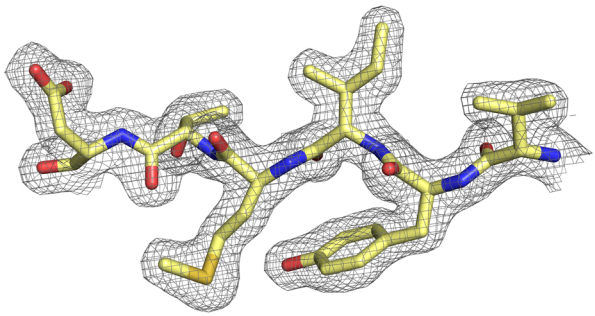

Figure 23: $2F_o$-F_c-electron density of CD1219 at sigma contour level 1.2.

7.2.3. SAXS analysis of CD1219

The oligomeric state of CD1219 in solution was analyzed by SAXS experiments. Data of native CD1219 were collected at BM29 ESRF (Grenoble, France; Pernot *et al.*, 2013). Best data and scattering curves were obtained for 5 mg/ml CD1219 in 20 mM Tris/HCl pH 8.0, 150 mM NaCl. After data averaging and buffer subtraction, a SAXS envelope was calculated using *DAMMIF* and the oligomeric state of CD1219 was analyzed using the program *Crysol*. The experimental SAXS data were fitted using theoretically calculated data for monomeric and dimeric CD1219 (Fig. 24A and B). The fit of experimental data with the theoretical curve for monomeric CD1219 can be described by a *X*-value of 217.8, the fit with dimeric CD1219 shows a *X*-value of 8.1. As a lower *X*-value indicates a better fit of experimental and theoretical data, SAXS analysis of CD1219 revealed that the protein is present as a dimer in solution. These results are in line with size exclusion chromatography experiments. The crystal structure of dimeric CD1219 was fitted into the calculated SAXS envelope (Fig. 24C) and the calculated SAXS envelope represents the overall shape of the protein in solution. As almost all parts of the CD1219 crystal structure are covered by the SAXS envelope, it can be assumed that the structure of CD1219 is identical in solution and in the crystal.

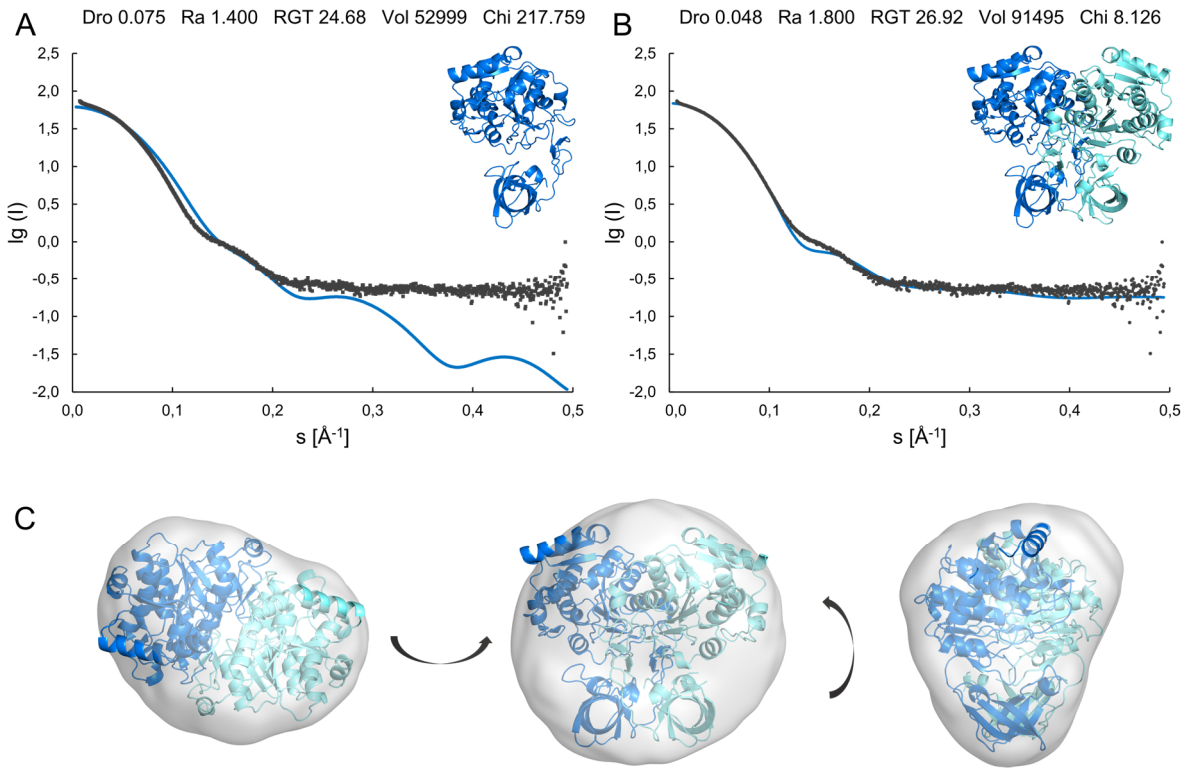

Figure 24: SAXS analysis of CD1219. (A) and (B) Experimental SAXS data (black dots) were fitted against theoretically calculated curves (blue line) for monomeric (A) and dimeric (B) CD1219. (C) The crystal structure of dimeric CD1219 was fitted into the calculated SAXS envelope and is shown in different orientations to illustrate the shape of CD1219 in solution.

7.2.4. Crystal structure and domain organization of dimeric CD1219

CD1219 is dimeric in solution (as shown by size exclusion chromatography and SAXS experiments). In the crystal state, two CD1219 molecules are present in the asymmetric unit, indicated by a Matthews coefficient of 2.58 and a solvent content of 52.26%. The crystal structure of dimeric CD1219 is depicted in Fig. 25A. The two chains are highlighted in marine and cyan, respectively. The dimer shows a butterfly shape and the monomer can be divided in a small N-terminal and a larger C-terminal domain that are connected via a small linker. The electron density for all amino acids of the CD1219 sequence was resolved in chain A, while density for amino acids 60-63 is missing in chain B. This indicates flexibility of this part of the protein.

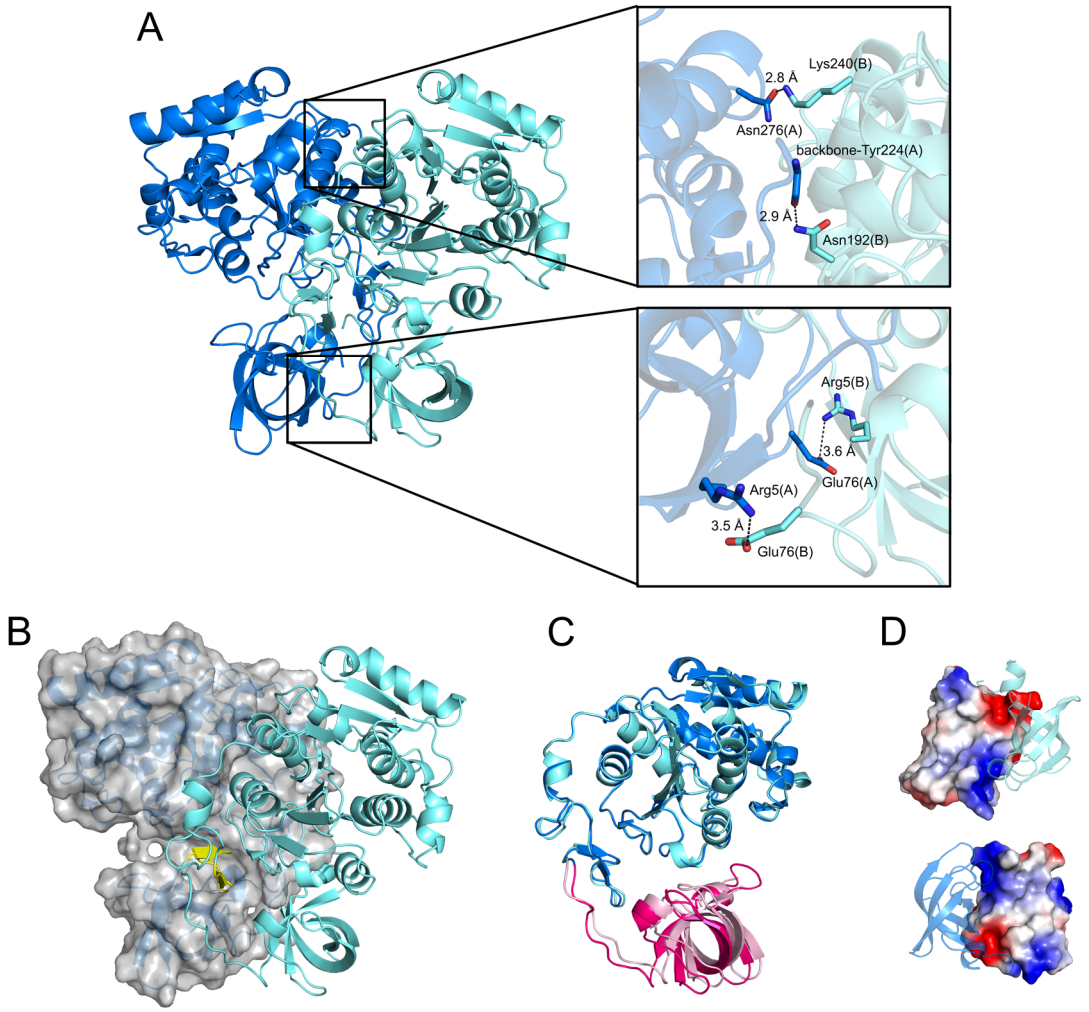

Figure 25: Crystal structure of dimeric CD1219 of *Clostridium difficile*. The dimer is shown in cartoon representation with the two chains in marine blue and cyan, respectively. (A) Exemplary interactions that stabilize the dimer interface. In the top right figure, two exemplary hydrogen bonds between Asn276 of chain A and Lys240 of chain B, as well as between the backbone oxygen of Tyr224 of chain A and Asn192 of chain B are shown. In the bottom right figure, the two salt bridges between Arg5 and Glu76 of each chain are depicted. (B) The two CD1219 chains interact tightly with each other. The β-strands (colored in yellow) that connect the large and the small domain of one chain insert into the groove between the large and small domain of the other chain and the other way around. This creates a handshake-like interaction between the two monomers. (C) The two CD1219 chains share the same secondary structure, while the small domains (colored in pink) are slightly shifted by 6.5° relative to each other. (D) Interaction of the small domains in the CD1219 dimer is stabilized by attractive electrostatic charge distribution.

The interaction surface of the two monomers is 3687.4 Å^2. The surface area of chain A and B is 16618 Å^2 and 16942 Å^2, respectively. The solvation free energy Δ^iG that is gained upon dimer formation is -46.3 kcal/mol. The dimer is stabilized by several hydrogen bonds and salt bridges between the 100 interface residues of chain A and the 94 interface residues of chain B. No disulfide bond is involved in the CD1219 dimer formation. Exemplary salt bridges and hydrogen bonds that stabilize the CD1219 dimer are depicted in Fig. 25A and a summary of all bonds involved in dimer formation is listed in table 21 (data generated by PISA server, Krissinel & Henrick, 2007). The dimer is further stabilized by the tight packing of both chains. The two β-strands that are located in the linker region between the small and the large domain

of one chain insert into the groove between the large and the small domain of the other chain and the other way around (Fig. 25B). This creates a handshake-like interaction of both monomers.

Both monomers can be aligned on top of each other with an r.m.s.d. of 1.44 Å including 352 of 355 residues. The large domains of chain A and B can be aligned with an r.m.s.d. of 0.57 Å and the small domains with an r.m.s.d. of 2.05 Å. The overall architecture of both monomers is the same, while the small domain of chain B is slightly rotated by 6.5° relative to the small domain of chain A (Fig. 25C). Due to the symmetry of the CD1219 dimer, the surfaces of the small domains that face each other show an opposite electrostatic charge distribution. The CD1219 dimer is hence further stabilized due to electrostatic attraction between the small domains of each chain (Fig. 25D).

Table 21: Overview of interactions between interface residues of chain A and chain B of the CD1219 dimer that stabilize the dimer formation. The residues of the respective chains are described by the three-letter amino acid code and their position in the CD1219 sequence. The atom involved in the salt bridge or hydrogen bond is indicated in brackets according to the typical PDB-file format. The bond distance is denoted in Å.

Salt bridges

No.	chain B residue	Dist. [Å]	chain A residue
1	Arg5 [NH1]	3.65	Glu76 [OE1]
2	Glu76 [OE1]	3.49	Arg5 [NH1]

Hydrogen bonds

No.	chain B residue	Dist. [Å]	chain A residue	No.	chain B residue	Dist. [Å]	chain A residue
1	His0 [H2]	1.65	Thr80 [OG1]	21	His0 [O]	2.10	Thr80 [H]
2	His0 [ND1]	3.85	Thr78 [O]	22	Met1 [O]	1.96	Lys96 [H]
3	Met1 [H]	2.16	Lys38 [O]	23	Lys4 [O]	2.30	Ser77 [H]
4	Lys4 [H]	2.19	Ser77 [O]	24	Ile39 [O]	1.95	Lys4 [HZ2]
5	Lys4 [HZ2]	2.02	Ser77 [OG]	25	Glu76 [OE2]	2.16	Val6 [H]
6	Lys4 [HZ1]	2.06	Ile39 [O]	26	Ser77 [O]	2.14	Lys4 [H]
7	Arg 5 [HH22]	2.20	Thr78 [OG1]	27	Ser77 [OG]	2.07	Lys4 [HZ3]
8	Val6 [H]	2.15	Glu76 [OE2]	28	Thr78 [OG1]	2.03	Arg5 [HH22]
9	Tyr36 [HH]	1.97	Tyr90 [O]	29	Arg89 [O]	2.16	Leu157 [H]
10	Asn52 [HD21]	2.06	Gln94 [O]	30	Tyr90 [O]	1.84	Tyr36 [HH]
11	Asn52 [HD22]	2.03	Pro92 [O]	31	Pro92 [O]	2.05	Asn52 [HD22]
12	Ser77 [H]	2.43	Lys4 [O]	32	Gln94 [O]	2.12	Asn52 [HD21]
13	Leu79 [H]	2.08	Ser3 [OG]	33	Gln94 [O]	3.89	Thr54 [OG1]
14	Thr80 [H]	2.32	His0 [O]	34	Asp153 [O]	1.95	Arg89 [HH22]
15	Arg89 [HH12]	1.90	Ile217 [O]	35	Ala155 [O]	2.19	Arg89 [HH21]
16	Lys96 [H]	2.09	Met1 [O]	36	Gly219 [O]	2.31	Gln103 [HE22]
17	Leu157 [H]	2.20	Arg89 [O]	37	Tyr224 [O]	1.97	Asn192 [HD21]
18	Asn192 [HD21]	2.06	Tyr224 [O]	38	Glu273 [O]	1.81	Lys240 [HZ2]
19	Lys240 [HZ1]	1.91	Asn276 [OD1]	39	Asn276 [OD1]	1.96	Lys240 [HZ1]
20	Lys240 [HZ2]	1.99	Glu273 [O]				

The CD1219 monomer consists of two domains. The small N-terminal domain (aa 1-72) forms a β-barrel that is made up of five antiparallel β-strands. It is flanked by two short α-helices that cap the top and bottom part of the β-barrel. A short linker (48 aa), which consists of two short antiparallel β-strands and a short α-helix, connects the small N-terminal domain with the larger C-terminal domain. This large C-terminal domain (aa 121-353) is characterized by a central β-sheet of eight parallel β-strands that are flanked by several α-helices. The domain organization of CD1219 is illustrated in Fig. 26C with the small N-terminal domain in pink, the short linker in yellow and the large C-terminal domain in blue. The colors of the respective domains will be retained during the following chapters. The structures of the small and the large domain are shown in Fig. 26A and B with the corresponding topology diagrams, respectively.

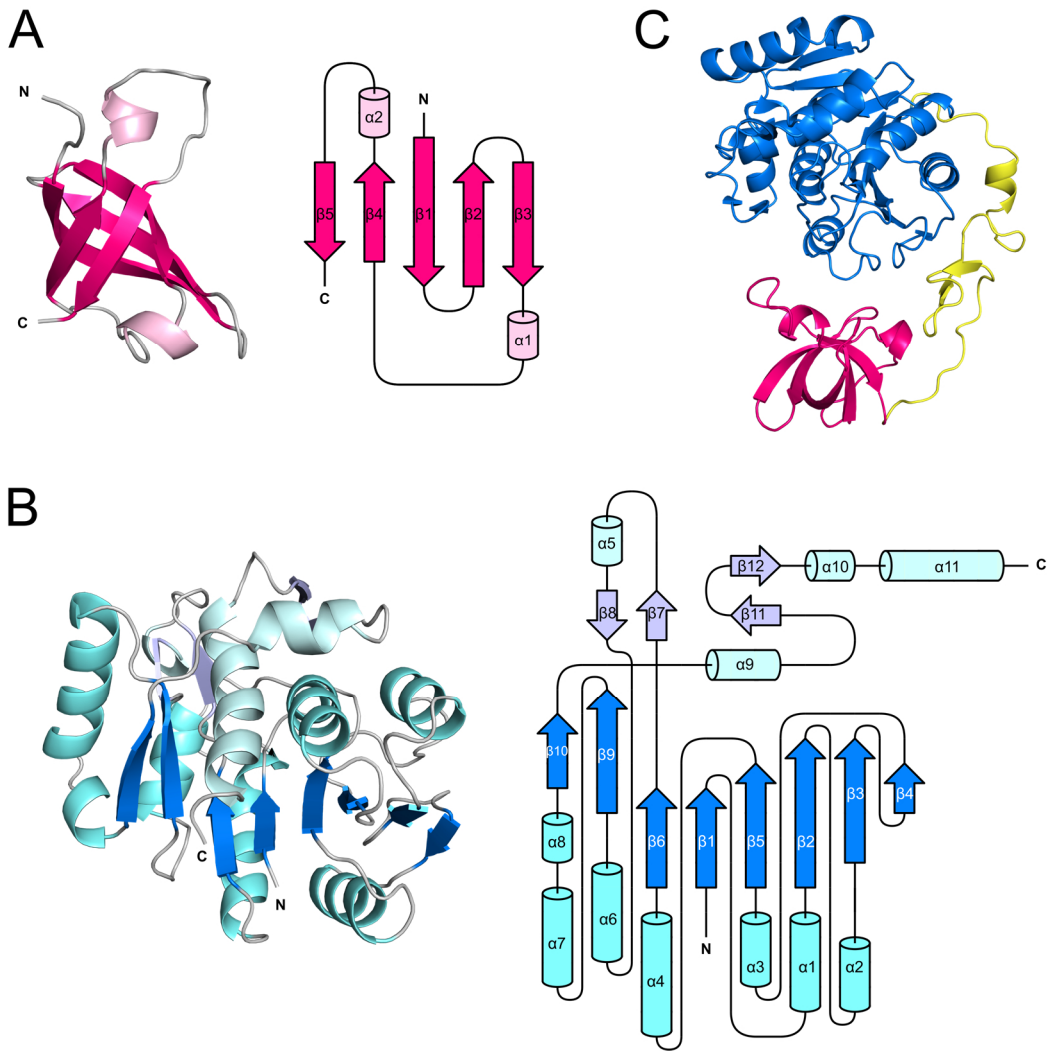

Figure 26: Domain organization of monomeric CD1219. (A) The small domain of CD1219 is shown in cartoon representation (left) and as topology diagram (right). The central β-barrel is colored in pink with the flanking α-helices in light pink. In the topology diagram, α-helices are shown as cylinders and β-sheets as arrows. The secondary structure motifs are numbered consecutively. (B) The large domain of CD1219 is illustrated in cartoon representation (left) and as two-dimensional topology diagram (right). The central β-strands are colored in blue and surrounding secondary structure motifs in light cyan. (C) Monomeric CD1219 consists of a small N-terminal domain (pink), followed by a linker (yellow) and a large C-terminal domain (blue).

7.2.5. Comparison of dimeric CD1219 with homologous structures

To get an idea about the function of CD1219, its crystal structure was compared to structurally similar proteins. For this purpose, the protein structure alignment servers DALI (Holm, *et al.*, 2016) and PDBeFold (Krissinel & Henrick, 2004) were used. A structural alignment is usually characterized by the r.m.s.d. and the sequence identity of the aligned protein structures. The PDBeFold Q-score is an indication for the alignment quality, as it includes both the r.m.s.d. value of the structural alignment as well as the sequence identity of the aligned proteins. A high Q-score indicates a promising structural alignment, while a value of 1 reveals identical protein structures. The DALI server uses the so-called Z-score for evaluation of the structural alignment. This score is similar to the PDBeFold Q-score, as it also indicates the order of structural similarity between the aligned proteins. It ranges between 0 and 10 with 10 showing the highest structural similarity. A Z-score higher than 2 usually indicates that the two aligned proteins are significantly similar, illustrating that they share a similar protein fold.
The search for structural homologous proteins of CD1219 was performed with dimeric CD1219, monomeric CD1219 and with the individual domains of CD1219, respectively. No structurally homologous proteins were found for the CD1219 dimer and monomer, but for the large and the small domain, respectively. The structures of the large and the small domain seem to be similar to conserved protein folds, but the combination of both domains in the CD1219 monomer and dimer might be unique. An analysis of the structural alignment results of the individual CD1219 domains can be found in 7.2.6 and 7.2.7.

7.2.6. The small domain of CD1219

The small N-terminal domain of CD1219 forms a five-stranded β-barrel with a hydrophobic core including Val9, Ile22, Val24, Ile43, Val49, Leu51 and Val68 (Fig. 27C). Strands β1-β4 form the common Greek key motif with three antiparallel β-strands and the fourth β-strand antiparallel and adjacent to the first one (Hutchinson & Thornton, 1993). One side of the small domain shows a neutral electrostatic surface charge with small positively charged regions that are created by the positively charged amino acids Lys4, Arg32 and Lys38 (Fig. 27A). This side of the small domain is directed to the cleft between the small and large domain and is hence not directly accessible. In addition, the aromatic amino acids of the small domain (Tyr34, Tyr36, Tyr65 and Phe67) are also accommodated on this side of the protein. On the opposite side of this domain, the surface shows a predominantly negative charge due to the presence of negatively charged amino acids Glu10, Ser11, Asn25, Glu29, Asn44, Asp47, Glu48 and Asn71 (Fig. 27B). This negatively charged surface of the small domain faces away from the rest of the CD1219 molecule and is hence accessible to potential binding partners.

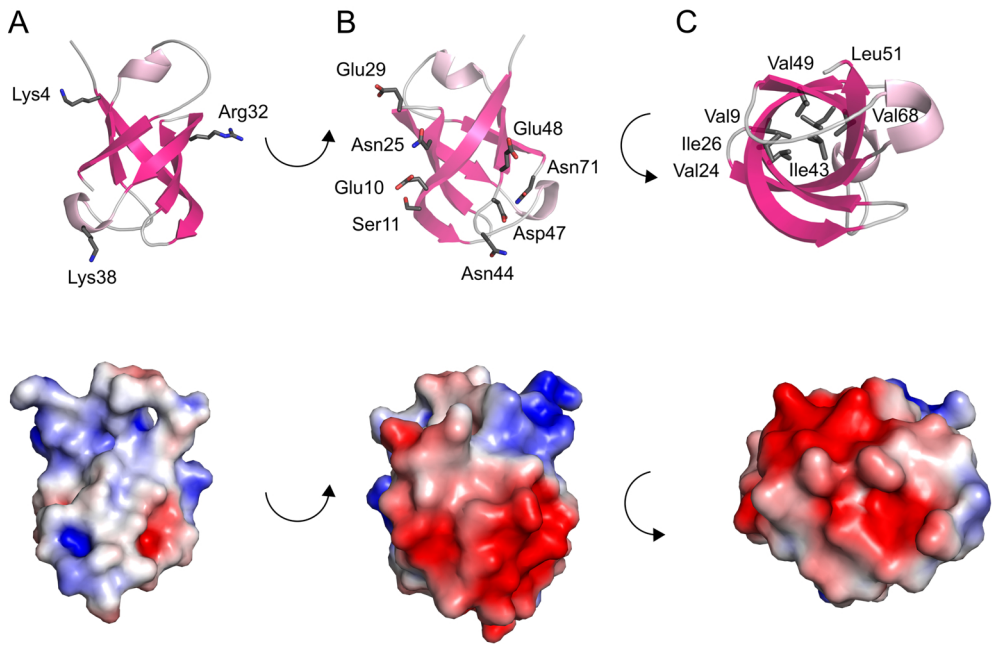

Figure 27: Amino acid distribution and electrostatic surface charge of the small N-terminal domain of CD1219. The crystal structure of the N-terminal domain of CD1219 is shown in cartoon representation with amino acid side chains as grey sticks. Representation (B) is a 180° turned representation of (A) and the orientation in (C) depicts the view from the top onto the β-barrel. The bottom panel of each figure illustrates the electrostatic surface charge distribution of the respective cartoon representation in the panel above.

The crystal structure of the small N-terminal domain of CD1219 was submitted to the structural alignment servers DALI (Holm & Laakso, 2016) and PDBeFold (Krissinel & Henrick, 2004) analogous to chapter 7.2.5. The results of this structural homology search are summarized in table 22.

The small domain of CD1219 shows an OB-fold (oligonucleotide/oligosaccharide binding), which is common among DNA-/RNA-binding proteins (Theobald *et al.*, 2003). The structural homology search (table 22) revealed a high similarity between the small N-terminal domain of CD1219 and DNA-/RNA-binding proteins, e.g. IF1s and Csps. Despite a low sequence similarity among OB-fold members, they all share a common protein fold characterized by a central β-barrel. The loops, which connect the central β-strands, are variable in their length and can contain further small secondary structure elements. In most cases, they are between 70 and 150 amino acids long and they can serve as “recognition domains” in larger proteins (Flynn & Zou, 2010). Several residues are conserved among OB-fold proteins, e.g. a glycine residue in the first half of β1 and a further glycine residue at the beginning of β4, probably mediating the termination of the preceding α-helix. In most cases, the ligand-binding site is located on strands β2 and β3 and can be additionally increased by the neighboring loops between β1-β2, β3-α, α-β4 and β4-β5 (Theobald *et al.*, 2003).

Table 22: Summary of the structural alignment search of the small domain of CD1219 using PDBeFold and DALI. Letters in brackets behind the PDB-Code indicate the protein chain that was used for the structural alignment. Proteins are listed with decreasing Q- and Z-score, respectively.

PDBeFold

PDB-Code	Protein (Organism)	Q-Score	R.m.s.d. (Å)	Sequence identity (%)
4QL5(A)	IF-1 (*Streptococcus pneumonia*)	0.542	1.5	15
3VYS(A)	HypC (*Thermococcus kodakarensis*)	0.540	1.4	20
5LMR(W)	IF-1 (*Thermus thermophilus*)	0.510	1.7	22
2N8N(A)	IF-1 (*Staphylococcus aureus*)	0.496	1.5	17
3I4O(B)	IF-1 (*Mycobacterium tuberculosis*)	0.482	1.5	18
5JX4(A)	Csp (*Bacillus caldolyticus*)	0.454	2.3	12
2LXJ(A)	Csp (*Listeria monocytogenes*)	0.445	2.3	14
2N3S(A)	IF-1 (*Burkholderia thailandensis*)	0.439	2.5	16
1CSQ(A)	Csp (*Bacillus subtilis*)	0.438	2.3	14
4QQB(Y)	Unr (*Drosophila melanogaster*)	0.436	2.1	17
2I5M(X)	Csp variant (*Bacillus subtilis*)	0.432	2.2	14
1HZB(A)	Csp (*Bacillus caldolyticus*)	0.429	2.3	17
3H43(C)	Nucleotidase (*Methanocaldococcus jannaschii*)	0.411	1.9	20
1MJC(A)	CspA (*Escherichia coli*)	0.405	2.8	15
1GUT(A)	MopII (*Clostridium pasteurianum*)	0.404	2.2	13

DALI

PDB-Code	Protein (Organism)	Z-Score	R.m.s.d. (Å)	Sequence identity (%)
4QL5(A)	IF-1 (*Streptococcus pneumoniae*)	8.6	1.9	15
5NO4(Z)	ribosomal subunit (*Escherichia coli*)	8.1	2.2	21
1ZO1(W)	IF-1 (*Escherichia coli*)	7.9	2.2	21
2N8N(A)	IF-1 (*Staphylococcus aureus*)	7.9	2.2	18
3I4O(B)	IF-1 (*Mycobacterium tuberculosis*)	7.9	2.5	17
5LMP(W)	IF-1 (*Thermus thermophilus*)	7.8	2.2	21
2ID0(B)	Exoribonuclease 2 (*Escherichia coli*)	7.6	3.0	11
5UZ4(Z)	Small ribosomal subunit biogenesis GTPase RsgA (*Escherichia coli*)	7.5	2.3	20
2WG6(C)	Transcription hydrolase (*Archaeoglobus fulgidus*)	7.5	2.0	18
3J80(i)	IF-1 (*Saccharomyces cerevisiae*)	7.3	2.3	10
1BKB(A)	IF-1 (*Pyrobaculum aerophilum*)	7.2	2.0	16
2RCN(A)	Probable GTPase EngC (*Salmonella enterica*)	7.2	1.5	23
3H43(C)	Nucleotidase (*Methanocaldococcus jannaschii*)	7.1	2.0	19
5GJQ(L)	Proteasome subunit (*Homo sapiens*)	7.1	1.8	18
3FP9(C)	ATPase (*Mycobacterium tuberculosis*)	7.0	2.0	21

Depending on their function and binding mode, OB-fold proteins can be grouped into different categories: proteins that bind specific regions of single-stranded nucleic acids, proteins that bind nucleic acids without strong sequence-specificity and proteins that bind non-helical nucleic acids. The function of OB-fold proteins is diverse, as they are involved in various processes, e.g. DNA replication, transcription, translation, cold shock response, etc., but most of them bind single stranded nucleic acids (ssDNA/ssRNA) (Arcus, 2002).

The structural homology search of the small domain of CD1219 revealed striking structural similarity to IF1s. IF1 is the smallest of the three translation initiation factors (IF1, IF2, IF3) that are essential for the formation of the translation initiation complex. They ensure the correct assembly and enhance the specificity of the complex (Hatzopoulos *et al*., 2010). IF1 binds to the A-site of the 30S subunit of the ribosome and guarantees the selection of the correct start codon. A crystal structure of IF1 from *Thermus thermophilus* in complex with the 30S ribosome (PDB-ID 1HR0; Carter *et al*., 2001) showed that IF1 binds to a cleft on the surface of the 30S subunit. Upon hydrogen bonding of IF1 and the ribosomal backbone, the RNA undergoes striking conformational changes. Two adenine residues are flipped out and bind to conserved arginine residues of IF1. This interaction is stabilized by further conserved IF1 residues and several hydrogen bonds and electrostatic interactions in this region (Carter *et al*., 2001).

To illustrate the similarities between the small domain of CD1219 and IF1s, a structure-based sequence alignment with different bacterial IF1s is shown in Fig. 28A. It reveals that the overall structure and topology of IF1s and the small CD1219 domain are highly similar. The only exception is the presence of the small helix α2 between strands β4 and β5 of CD1219 that can't be found in the selected IF1s and an extended loop between β1 and β2 of CD1219. To illustrate these structural similarities and differences, a structural alignment of the small N-terminal domain of CD1219 from *Clostridium difficile* and full-length IF1 from *Streptococcus pneumoniae* (PDB-ID 4QL5; Stogios *et al*., 2014, unpublished) is shown in Fig. 28B. Both structures can be aligned with an r.m.s.d. of 2.05 Å over 59 of 71 residues and share a sequence identity of 19.3%.

The structure-based sequence alignment illustrates that several residues of CD1219 and IF1s, e.g. a glycine residue at the beginning of β1, β3 and β4, are conserved. These glycine residues can induce a break in the secondary structure motifs, as they are usually located at the end or beginning of a β-strand or α-helix. However, the residues of IF1 that are involved in nucleotide binding, e.g. Lys39, Arg41, Arg46 (marked in red in Fig. 28A) are not conserved in CD1219. This indicates that, although CD1219 and IF1s share an OB-fold, they might use a different ligand, a different way of ligand binding or might possess a different function at all.

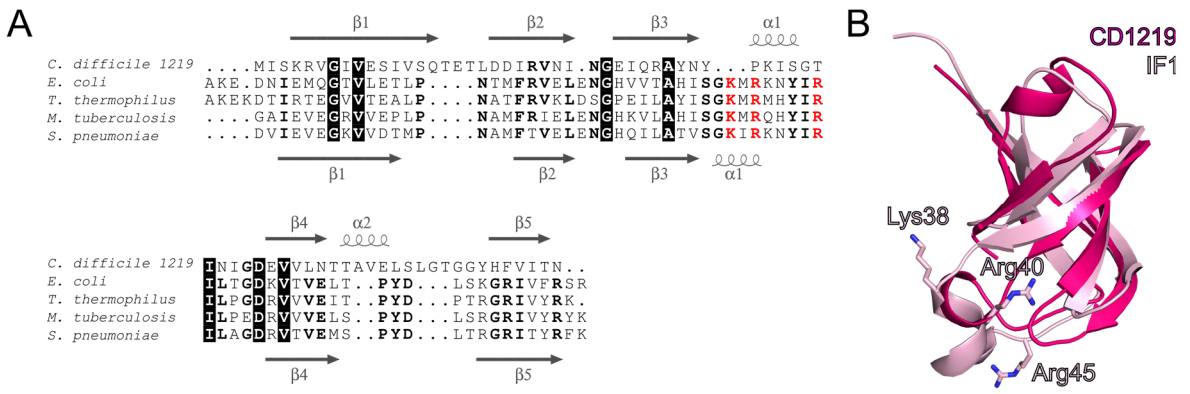

Figure 28: (A) Structure-based sequence alignment of the small N-terminal domain of CD1219 from *Clostridium difficile* and bacterial IF-1s from *Escherichia coli* (PDB-ID 1AH9; Sette *et al.*, 1997), *Thermus thermophilus* (PDB-ID 1HR0; Carter *et al.*, 2001), *Mycobacterium tuberculosis* (PDB-ID 3I4O; Hatzopoulos & Mueller-Dieckmann, 2010) and *Streptococcus pneumoniae* (PDB-ID 4QL5; Stogios *et al.*, 2014, unpublished). Secondary structure motifs of CD1219 from *Clostridium difficile* and IF-1 from *Streptococcus pneumoniae* are depicted above and under the sequence alignment, respectively. β-strands are illustrated as grey arrows, α-helices are shown as grey spirals. The secondary structure motifs are numbered consecutively. Conserved amino acids are indicated by bold letters. Amino acids that are conserved in all shown bacterial species are highlighted by black boxes. Amino acids of IF1 that are involved in RNA binding are shown in red. (B) Structural alignment of the N-terminal domain of CD1219 from *Clostridium difficile* (pink) and IF-1 from *Streptococcus pneumonia* (light pink). Both structures are shown in cartoon representation. Residues involved in DNA binding of IF1 are shown as sticks.

To map the RNA-binding mode of IF1s on CD1219, the RNA molecule bound to IF1 of *Thermus thermophilus* (PDB-ID 1HR0; Carter *et al.*, 2001) was aligned onto the structure of the small domain of CD1219 (Fig. 29). As illustrated by the structure-based sequence alignment (Fig. 28A), the residues of IF1 (Lys39, Arg41, Arg46) that are involved in RNA binding are not present in the small domain of CD1219. Furthermore, the RNA molecule clashes with some regions of CD1219. A comparison of the electrostatic surface of the RNA binding region of IF1 and CD1219 also reveals differences between both proteins. The RNA binding surface of IF1 is positively charged (Fig. 29C), whereas the surface of the small domain of CD1219 is predominantly negatively charged (Fig. 27). Although the part of CD1219 that was aligned onto the positively charged RNA binding surface of IF1 shows small positively charged patches, these patches are located in the region that points away from the solvent environment. These differences between IF1 and CD1219 indicate that either other parts of the small domain of CD1219 are involved in nucleic acid binding or conformational changes of CD1219 have to take place upon potential DNA-/RNA binding.

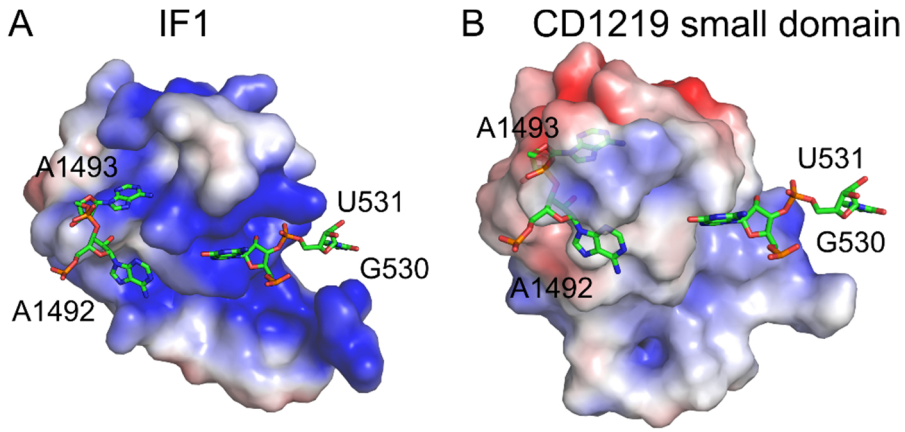

Figure 29: (A) Electrostatic surface of IF1 (PDB-ID 1HR0; Carter *et al.*, 2001) bound to 30S ribosomal RNA from *Thermus thermophilus*. Residues that are involved in RNA binding are shown as sticks. (B) The 30S ribosomal RNA from *Thermus thermophilus* was aligned to the small domain of CD1219 from *Clostridium difficile*. The surface of the small domain of CD1219 in the same orientation of IF1 in (A) was colored according to its electrostatic charge distribution (calculated with APBS) and the aligned residues of the 30S RNA from *Thermus thermophilus* are shown as sticks.

Besides IF1s, the small N-terminal domain of CD1219 shows striking structural similarity to “cold-shock” proteins (table 22). These proteins are expressed upon temperature decrease and enable bacteria to survive under stress conditions. Under these conditions, the rate of transcription and translation is downregulated and enzyme activity, proper protein folding and ribosome assembly is disturbed. Cold-shock proteins thereby serve as chaperons to counteract these harmful conditions and ensure translation initiation by destabilizing secondary structures of mRNA (Chaikam & Karlson, 2010). Csps are small (65-70 aa) proteins that bind single-stranded nucleic acids. They contain a so-called cold shock domain (CSD), which is highly conserved among Csps from all kingdoms and is involved in nucleic acid binding (Graumann & Marahiel, 1996). It contains the RNA binding sequence motifs RNP1 and RNP2 (ribonucleoprotein motifs) (Schindelin *et al.*, 1993), which are mainly composed of basic and aromatic residues located primarily in β-strands β1 and β2. These motifs can bind regions of six to seven nucleotides of single-stranded nucleic acids with high specificity. Although the sequence and structure conservation among bacterial Csps is extremely high, they show different thermal stabilities based on the electrostatic surface charge distribution (Carvajal *et al.*, 2017).

The structural alignment of the small domain of CD1219 and a Csp from *Bacillus caldolyticus* (PDB-ID 1CSJ, Bressanelli *et al.*, 1999) is exemplarily shown in Fig. 30B. Both structures can be aligned with an r.m.s.d. of 2.89 Å over 32 of 65 residues and show a sequence identity of 12%. They share the same topology with CD1219 possessing an additional short α-helix (α2) between strands β4 and β5. Binding of the Csp to single-stranded nucleic acids is mediated

by the two RNP motifs. They are indicated by red boxes in the sequence alignment and are conserved among a variety of RNA binding proteins. The RNP motifs are located in strands β2 and β3, respectively, and consist mainly of aromatic and basic residues, resulting in a positively charged electrostatic surface that enables binding to the negatively charged phosphate backbone of nucleic acids. The aromatic residues are essential for binding single-stranded nucleic acids, as they mediate the binding to nucleic acids by base stacking (Chaikam & Karlson, 2009). This was shown by mutational studies in *E. coli* Csp, where the two phenylalanine residues Phe16 and Phe58 were mutated to alanine residues. This mutation resulted in a dramatic reduction of the Csp binding affinity for ssRNA (Brennan & Platt, 1991). The structure-based sequence alignment in Fig. 30A shows that these Csp residues that are involved in nucleic-acid binding are not conserved in the small N-terminal domain of CD1219. The positive surface charge attributed to the aromatic and basic residues of Csp from *Bacillus caldolyticus* is concentrated on one side of the protein, indicating that this side is the region involved in binding to ss-nucleic acids. In contrast, the positive electrostatic surface charge of the small domain of CD1219 is distributed over a larger surface area (Fig. 31).

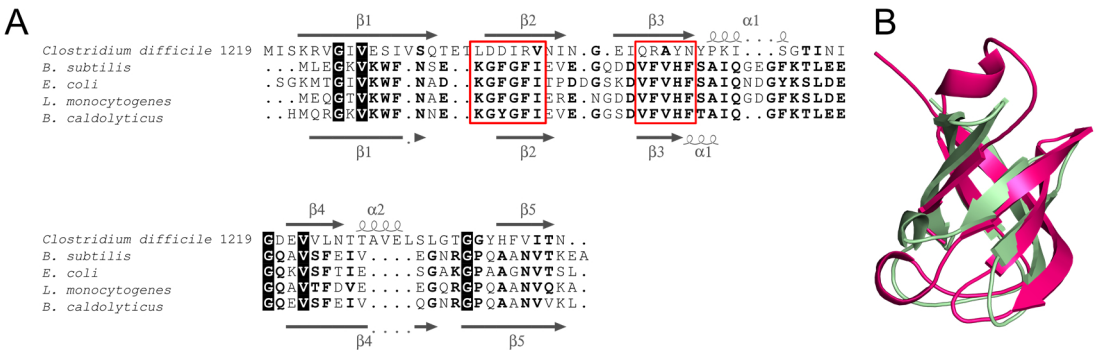

Figure 30: (A) Structure-based sequence alignment of the small N-terminal domain of CD1219 and bacterial Cold-shock proteins (Csp) from *Bacillus subtilis* (PDB-ID 1CSQ, Bressanelli *et al.*, 1999), *Escherichia coli* (PDB-ID 1MJC, Schindelin *et al.*, 1994), *Listeria monocytogenes* (PDB-ID 2LXJ, Lee *et al.*, 2013) and *Bacillus caldolyticus* (PDB-ID 5JX4, Carvajal *et al.*, 2017). Secondary structure motifs of CD1219 from *Clostridium difficile* and Csp from *Bacillus caldolyticus* are depicted above and under the sequence alignment, respectively. β-strands are illustrated as grey arrows, α-helices are shown as grey spirals. The secondary structure motifs are numbered consecutively. Conserved amino acids are indicated by bold letters. Amino acids that are conserved among all shown bacterial species are highlighted by black boxes. Red boxes indicate the RNA-binding motifs RNP1 and RNP2, respectively. (B) Exemplary structural alignment of the N-terminal domain of CD1219 from *Clostridium difficile* (pink) and Csp from *Bacillus caldolyticus* (light green). Both structures are shown in cartoon representation.

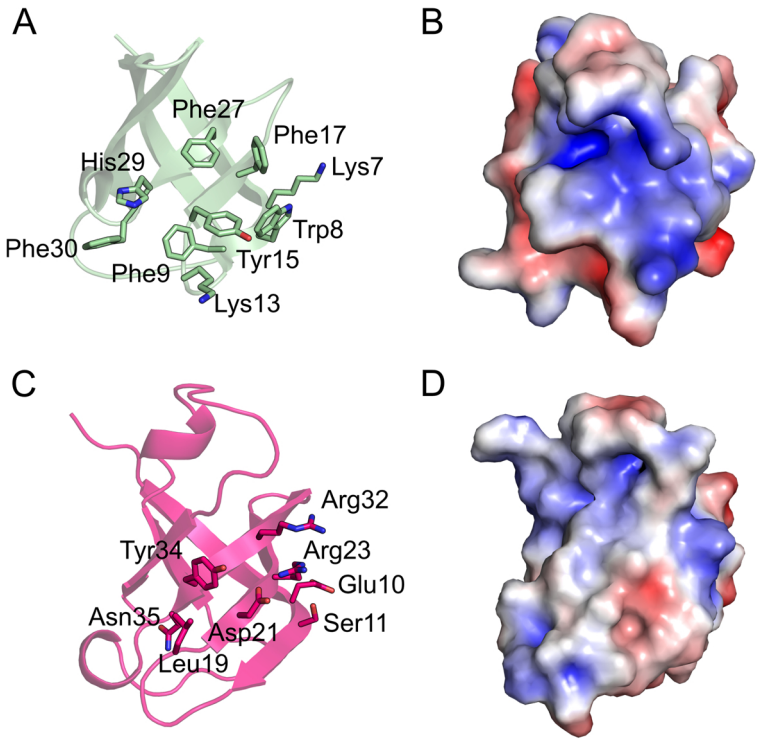

Figure 31: (A) Residues involved in single-stranded nucleic acid binding of Csp from *Bacillus caldolyticus* and respective aligned residues of the small domain of CD1219 from *Clostridium difficile* in (C). The Csp crystal structure (PDB-Code 5JX4) is shown in light green, the amino acid side chains that are involved in nucleic acid binding are shown as green sticks. The crystal structure of the small CD1219 domain is shown in pink. (B) The electrostatic surface representation of the Csp structure in the same orientation as the crystal structure in (A) was calculated with APBS (Baker *et al.*, 2001). (D) The electrostatic surface of the small domain of CD1219 of *Clostridium difficile* after alignment to Csp of *Bacillus caldolyticus* shows differences in the charge distribution of both proteins.

To sum up, despite low sequence identity, the small N-terminal domain of CD1219 from *Clostridium difficile* shows striking structural similarity to OB-fold nucleic-acid binding proteins such as IF1s and Csps. Although they share the same fold, CD1219 lacks the conserved residues that are responsible for DNA-/RNA-binding in IF1s and Csps. Based on the crystal structure of CD1219, a nucleic-acid binding region could not be clearly defined. The nucleic-acid binding of OB-fold proteins usually takes place in the β2-β3 region of the central β-barrel (Lee *et al.*, 2013). These β-strands are located at the outside of dimeric CD1219 and not in the dimer interface. Usually a positively charged surface is necessary to bind the negatively charged nucleic acid backbone. The β2-β3 region of the CD1219 OB-fold shows positively charged patches including lysine and arginine residues. These patches are not concentrated in one specific region of the small domain of CD1219 in contrast to IF1s and Csps. This indicates that either conformational changes would be necessary upon DNA-/RNA-binding to CD1219 or that CD1219 would use a binding mechanism that is different from that of IF1s and Csps, if the small domain of CD1219 is indeed able to bind nucleic acids.

The above-mentioned similarities between the small domain of CD1219 and OB-fold proteins allow a construction of a hypothetical nucleic acid binding model. The nucleic acid could either bind separately on both sides to the small domains of the CD1219 dimer or it could bind to both small domains simultaneously and thus wrap around the CD1219 dimer (Fig. 32). It was shown that several OB-fold proteins undergo conformational changes upon nucleic acid binding. Furthermore, it is possible that the nucleic acid itself changes its conformation upon binding to OB-fold proteins, as was shown for IF1s. This conformational change could be necessary in nature to distinguish between the bound and unbound form of the protein (Williamson, 2000). OB-fold proteins can be monomeric (e.g. IF1s), but can also be present as dimers (Csp from *Bacillus caldolyticus*, Carvajal *et al.*, 2017) or higher oligomers (nucleotidase from *Methanocaldococcus jannaschii*, Zhang *et al.*, 2003). In this case, dimerization of the nucleic acid-binding domain is often used to increase the affinity and/or specificity of single-stranded nucleic acid binding.

Due to its structural similarity to nucleic acid-binding proteins and its structural arrangement, the small N-terminal domain CD1219 could indeed function as a nucleotide-binding domain. As no hypotheses about the mechanism and substrate could be derived from the crystal structure of the small domain of CD1219, further experiments are necessary to prove the potential binding of nucleic acids to CD1219.

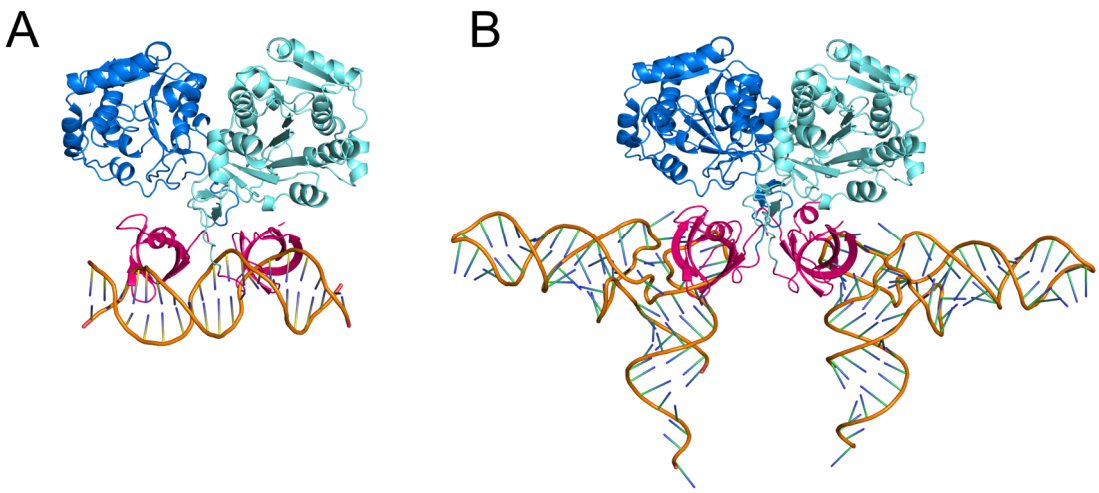

Figure 32: Potential models of nucleic acid binding to the small N-terminal domain of CD1219. The CD1219 dimer is shown in cartoon representation in blue with the small N-terminal domains in pink. The nucleic acid is shown as orange cartoon representation. One nucleic acid molecule can either bind to both small CD1219 domains simultaneously (A) or to each small domain separately (B).

Although the small domain of CD1219 shows an OB-fold, it does not necessarily have to be involved in nucleic acid binding. The largest subfamily of OB-fold proteins is the nucleic-acid binding protein family, but in contrast to this some OB-fold proteins are known that mediate interaction of proteins or assembly of protein complexes (Theobald *et al.*, 2003; Lynn *et al.*, 2010). A recently published structure illustrates that the cell-puncturing complex of

bacteriophage T4 also includes an OB-fold protein (Arcus, 2002). The OB-fold part probably mediates the contact between the individual protein components of this complex. To elucidate the exact molecular function of the small domain of CD1219, whether it is involved in nucleic acid binding or protein-protein interaction, further experiments are essential.

7.2.7. The large domain of CD1219 as a potential GTPase

The crystal structure and the corresponding topology diagram of the large C-terminal domain of CD1219 of *Clostridium difficile* is shown in Fig. 26B. It shows an equal distribution of negative and positive charge on its surface, while the solvent-exposed part of this domain is predominantly negatively charged (Fig. 33).

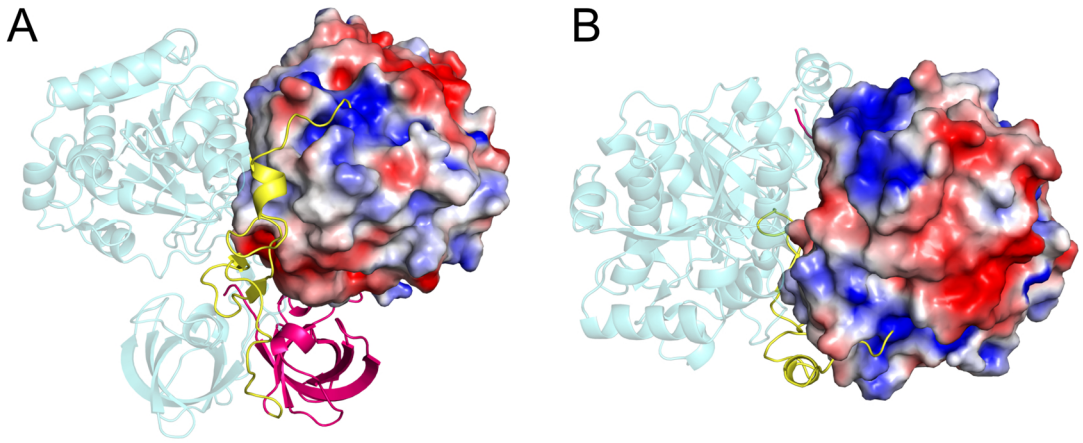

Figure 33: Electrostatic surface of the large C-terminal domain of CD1219. The dimer of CD1219 is shown in cartoon representation with the large domain of one chain colored according to its electrostatic charge distribution. The surface of the large domain is shown from the top in (B).

Analogous to the small N-terminal domain (chapter 7.2.6), a structural homology search was performed with the large C-terminal domain of CD1219 from *Clostridium difficile* using DALI and PDBeFold. The results are summarized in table 23. Among the proteins that showed the best structural alignment to the large domain of CD1219 are mainly P-loop NTPases that belong to the SIMIBI class of NTPases (Leipe *et al.*, 2002). The best structural alignment of the large domain of CD1219 was obtained for the plasmid partitioning protein ParA from *Escherichia coli* (PDB-ID 5U1G, Zhang & Schumacher, 2017), the signal recognition particle receptor FtsY from *Arabidopsis thaliana* (PDB-ID 2OG2, Chandrasekar *et al.*, 2007) and the cell division protein MinD from *Archaeoglobus fulgidus* (PDB-ID 1HYQ, Cordell & Lowe, 2001) with Z-scores of 9.5, 8.6 and 7.8, respectively.

Table 23: Summary of the structural alignments of the large C-terminal domain of CD1219 from *Clostridium difficile* with homologous proteins using PDBeFold and DALI. Letters in brackets behind the PDB-Code indicate the protein chain that was used for the alignment. Proteins are ordered with decreasing Q- and Z-score, respectively.

PDBeFold					**DALI**				
PDB-Code	Protein (Organism)	Q-Score	R.m.s.d. (Å)	Sequence identity (%)	PDB-Code	Protein (Organism)	Z-Score	R.m.s.d. (Å)	Sequence identity (%)
2L82(A)	P-loop NTPase (synthetic construct)	0.24	2.72	10	4DZZ(B)	ParF (*Escherichia coli*)	9.6	3.2	12
5GAJ(A)	P-loop fold protein (synthetic construct)	0.23	2.54	13	5U1G(A)	ParA protein (unidentified)	9.5	3.2	12
5U1G(D)	ParA protein (unidentified)	0.21	3.50	10	2OG2(A)	FtsY (*Arabidopsis thaliana*)	8.6	3.3	15
5GAD(I)	Signal recognition particle receptor FtsY (*Escherichia coli*)	0.20	3.46	11	2J7P(E)	FtsY (*Thermus aquaticus*)	8.4	3.3	12
1HYQ(A)	Cell division regulator MinD (*Archaeoglobus fulgidus*)	0.20	2.79	8	3DM5(A)	Signal recognition protein (*Pyrococcus furiosus*)	8.3	3.5	11
2WSM(A)	HypB (*Archaeoglobus fulgidus*)	0.19	3.35	8	1VMA(A)	FtsY (*Thermotoga maritima*)	8.3	3.5	14
5J1J(B)	FleN (*Pseudomonas aeruginosa*)	0.19	3.43	11	2P67(A)	LAO/AO transport system kinase (*Escherichia coli*)	8.0	3.7	12
3R9J(B)	MinD (*Escherichia coli*)	0.19	2.97	11	2XXA(B)	Signal Recognition Particle (*Escherichia coli*)	8.0	3.4	15
4RFV(B)	Aps kinase (*Mycobacterium tuberculosis*)	0.19	2.81	8	4C7O(A)	Signal Recognition Particle (*Escherichia coli*)	8.0	3.3	9
3CWQ(B)	ParA (*Synechocystis sp.* PCC 6803)	0.19	3.69	7	1O87(B)	Ffh GTPase (*Thermus aquaticus*)	8.0	3.2	5
4PNH(H)	Phosphatase (*Burkholderia thailandensis*)	0.19	2.95	10	4LC1(B)	MeaB (*Methylobacterium extorquens*)	7.9	3.3	8
4RZ3(B)	MinD-like ATPase FlhG (*Geobacillus thermodenitrificans*)	0.19	3.13	12	2QM7(B)	MeaB (*Methylobacterium extorquens*)	7.9	3.1	8
1DTS(A)	Dethiobiotin Synthetase (*Escherichia coli*)	0.19	3.54	9	3R9J(B)	MinD (*Escherichia coli*)	7.8	3.3	15

P-loop GTPases share a conserved protein fold that is characterized by a central β-sheet composed of seven (mainly) parallel β-strands that are each flanked by an α-helix. They are characterized by several conserved residues and secondary structure motifs (Saraste *et al.*, 1990). To compare the topology of selected P-loop GTPases with the large domain of CD1219, the structures were aligned respective to the central β-sheet. The strands involved in the central β-sheet were numbered from 1-7 consecutively, discounting further interspersed β-strands that are not involved in the central β-sheet. A comparison of the topology and overall structure of some P-loop GTPases selected from table 23 and the large C-terminal domain of CD1219 is shown in Fig. 34. A structure-based sequence alignment of these selected P-loop GTPases and the large domain of CD1219 is shown in Fig. 35.

One prominent structural feature involved in nucleotide binding in all GTPases is the so-called P-loop, also called Walker A motif. It is usually located as a flexible loop between strand β1 and the following α-helix α1 (indicated as a red line in the topology diagrams in Fig. 34 and as a red box in the sequence alignment in Fig. 35). The typical P-loop amino acid sequence (GXXXXGK) is conserved among most GTPases and is characterized by several glycine residues (Saraste *et al.*, 1990). It is involved in correctly positioning the triphosphate part of the bound nucleotide. Another conserved feature of P-loop GTPases is the [NT]KXD motif, which determines the guanosine specificity of the GTPase (Bourne *et al.*, 1991).

The alignment of the large domain of CD1219 from *Clostridium difficile* with the selected P-loop GTPases shows that a flexible glycine-rich P-loop cannot be found between β1 and α1 of CD1219 (Fig. 34). However, a similar glycine-rich loop with the amino acid sequence GXGXXGXGXXXGXT is present between strand β5 and helix α4 of CD1219. In the CD1219 crystal structure, this loop is involved in the coordination of a bound phosphate molecule and could hence be described as a “P-loop” involved in potential nucleotide binding.

Another common feature of P-loop GTPases is the Walker B motif, usually located distal to the Walker A motif. The Walker B motif commonly involves a conserved aspartate or glutamate residue that is located at the end of a β-strand and coordinates a water-bridged magnesium ion (Walker *et al.*, 1982). This conserved aspartate residue (D153; marked in red in the sequence alignment in Fig. 35) can also be found in the large domain of CD1219 and is located at the end of strand β2. In contrast to most GTPases, this conserved aspartate residue is not involved in magnesium binding in the crystal structure of CD1219.

In general, the P-loop GTPase family can be subdivided in two groups: TRAFAC and SIMIBI GTPases (Leipe *et al.*, 2002). TRAFAC is short for tranlation factor-related class of GTPases and involves translation factors, heterotrimeric G proteins, dynamins etc. The SIMIBI class includes signal recognition particle-associated GTPases and various enzymes involved in metabolism. It was named after its three largest subgroups: signal recognition GTPases, the MinD and BioD (dethiobiotin synthetase) superfamily (Leipe *et al.*, 2002). The main structural

difference between TRAFAC and SIMIBI GTPases is the arrangement of the β-strands that form the central β-sheet. In SIMIBI GTPases, all β-strands are arranged parallel to each other. In TRAFAC GTPases, the β-strand flanking the Walker B-containing strand is antiparallel to the others. On basis of their amino acid sequence, both GTPase classes show differences concerning certain conserved residues. E.g., in the TRAFAC class, a highly conserved threonine or serine residue is present in the loop between β2 and β3, besides a conserved serine residue in β7 that is not conserved among SIMIBI GTPases.

The overall topology of the C-terminal domain of CD1219 from *Clostridium difficile* and P-loop SIMIBI GTPases is strikingly similar, as both contain a central seven-stranded parallel β-sheet with flanking α-helices. The main difference is the additional short β-sheet β3* in CD1219 between β3 and β4 that is not present in other SIMIBI GTPases. Further short α-helices and β-sheets in the loop regions following β5 and β7 are present in the CD1219 structure that can't be found in the other selected GTPases. Furthermore, the large domain of CD1219 lacks the above-mentioned conserved residues that are characteristic for the TRAFAC class of GTPases.

Moreover, the SIMIBI class is characterized by a further conserved aspartate residue at the beginning of β4 (part of the Walker B motif), which exact function is unknown. This aspartate residue is also present in the CD1219 structure and is indicated in red in the sequence alignment in Fig. 35. The P-loop sequence of SIMIBI GTPases includes a third conserved glycine residue (GXXGXGK), similar to the P-loop sequence of CD1219 (Leipe *et al.*, 2002). The only difference between CD1219 and traditional P-loop GTPases is the position of the P-loop. The [NT]KXD motif conserved among GTPases is generally located in the loop following β6 and shows strong variability in SIMIBI GTPases. This conserved motif is not present in the loop following β6 in CD1219, but a similar motif if located following strand β8 in the large domain of CD1219 (sequence alignment in Fig. 35). Another factor indicating that CD1219 might belong to the SIMIBI class of GTPases is the fact that it is dimeric in solution, as most SIMIBI GTPases are present as dimers in nature (Leipe *et al.*, 2002).

As the core structure of CD1219 is strikingly similar to P-loop SIMIBI GTPases and due to the presence of several conserved residues and structural motifs of P-loop SIMIBI GTPases, the large domain of CD1219 could most probably be a SIMIBI GTPase. As some of the structural motifs that are essential for GTP binding are not located in the same region as in other GTPases, the GTP binding region of CD1219 is probably located in a different region of the protein as in homologous traditional GTPases.

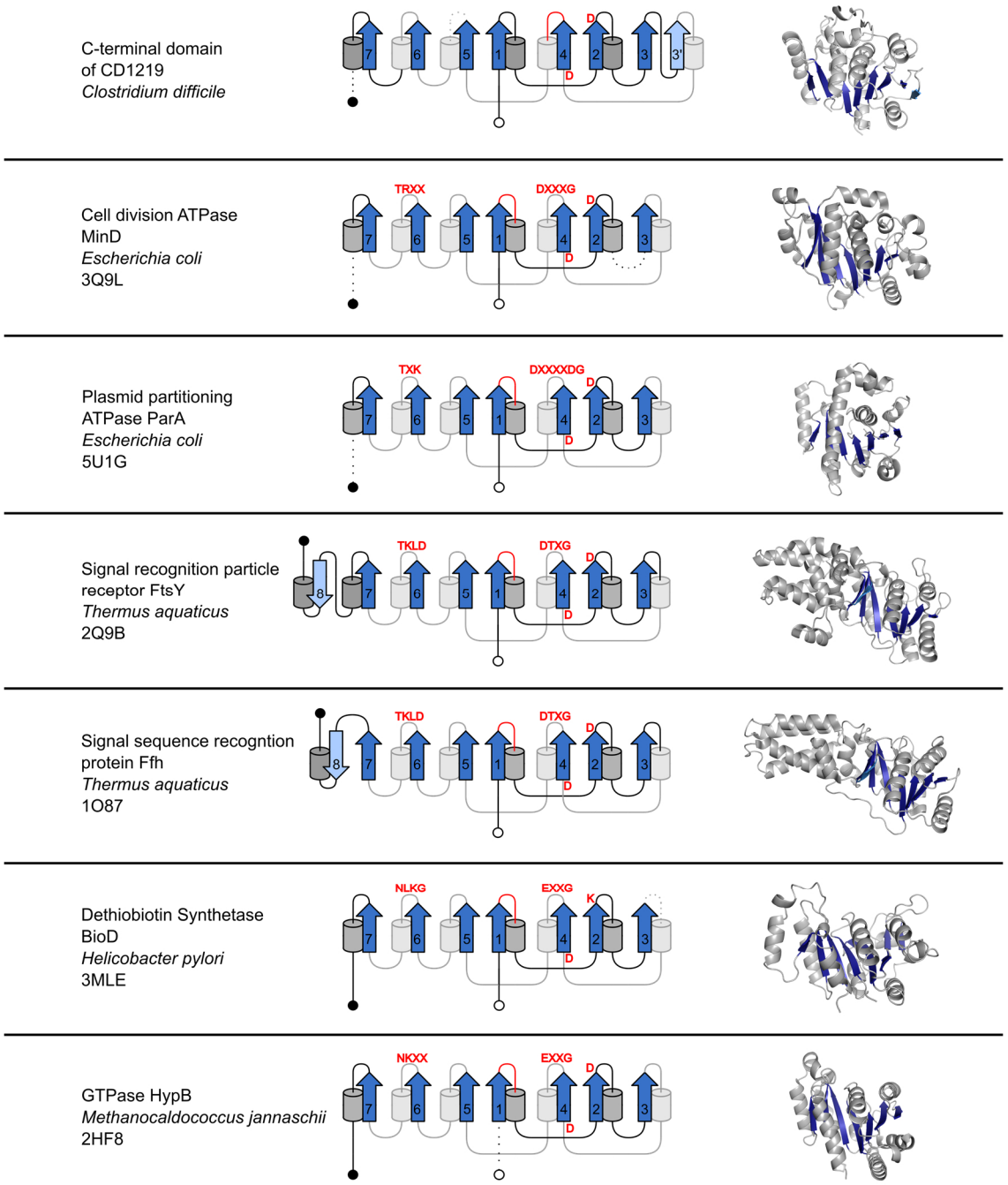

Figure 34: Comparison of selected P-loop NTPases. The topology diagrams and the overall structure of selected P-loop NTPases are shown. Selected proteins are described by annotated protein name, species name and PDB-Code in the left panel. The corresponding topology diagram is shown in the middle column. β-strands are shown as arrows with the arrowhead on the C-terminal side and are numbered from 1-8. The central β-sheet is colored in dark blue, additional β-sheets are highlighted in light blue. α-helices are shown as cylinders. Helices above the plane of the central β-sheet are colored in dark grey, helices below the β-sheet plane are illustrated in light grey. The N-terminus is indicated by a hollow circle, the C-terminus by a filled black circle. The P-loop is shown as a red line, further conserved structural motifs and residues are highlighted in red. Dashed lines indicate secondary structure motifs that were left out in the topology diagram for clarity.

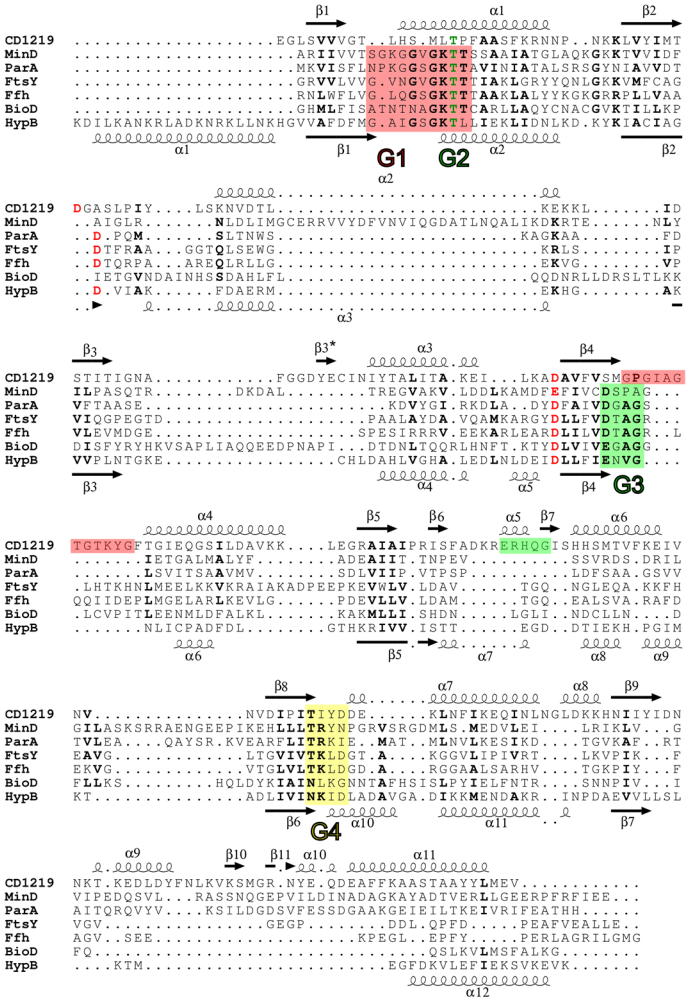

Figure 35: Structure-based sequence alignment of the large C-terminal domain of CD1219 from *Clostridium difficile* with selected SIMIBI GTPases. The alignment was created with PromalS3D and edited with Espript. Bold characters indicate conserved amino acid residues. Residues in red indicate conserved amino acids involved in nucleotide binding. The secondary structure of CD1219 is shown above the alignment with arrows indicating β-strands and spirals illustrating α-helices. The secondary structure of the exemplary GTPase HypB from *Methanocaldococcus jannaschii* is shown under the alignment. Red boxes indicate the conserved P-loop region, yellow and green boxes show the modified DXXD and TXXD motifs, respectively.

To study potential GTP-binding to the large domain of CD1219 of *Clostridium difficile*, the binding mode of GTP to the homologous GTPase HypB was analyzed. The binding of the substrate analogue GSP to the GTPase HypB from *Methanocaldococcus jannaschii* (PDB-ID 2HF8, Gasper *et al.*, 2006) is illustrated in Fig. 36A. The conserved secondary structure elements and amino acids that are involved in GSP binding are highlighted in respective colors analogous to the sequence alignment in Fig. 35. By aligning the large domain of CD1219 on HypB, it can be clearly shown that the bound GSP molecule clashes with secondary structure elements, especially helix α1, of CD1219 (Fig. 36B). Compared to HypB, α1 of CD1219 is longer than the aligned helix α1 in HypB. The clashes of the aligned GSP and CD1219 indicate that either the nucleotide-binding site of CD1219 is located in a different position as compared to HypB or that CD1219 has to undergo significant conformational changes upon nucleotide binding. The different positions of the P-loop and location of the conserved aspartate residue in HypB and CD1219 are illustrated in Fig. 36C.

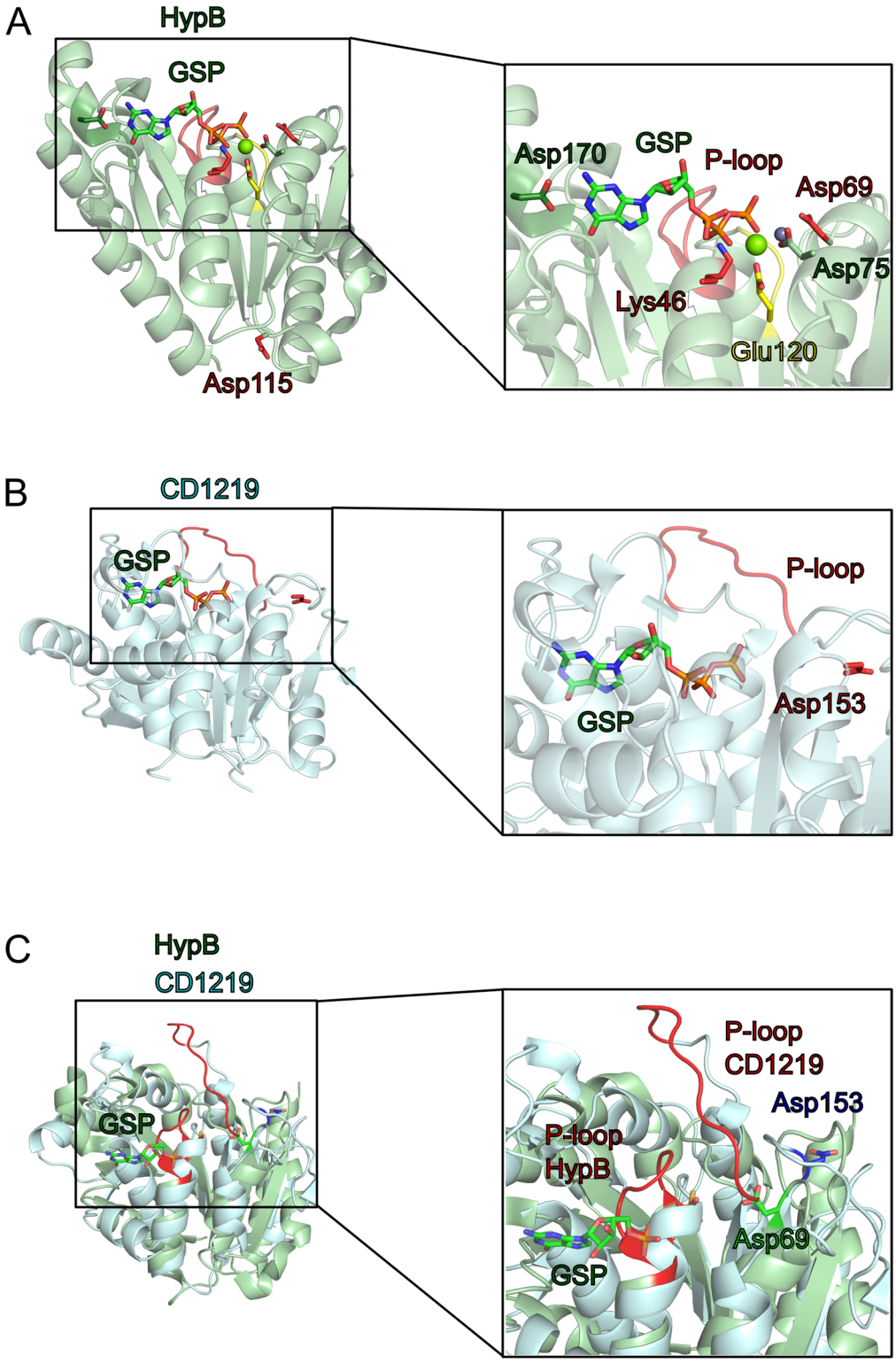

Figure 36: (A) GSP bound to HypB from *Methanocaldococcus jannaschii* (PDB-ID 2HF8, Gasper *et al.*, 2006). HypB is shown as cartoon representation in light green with the conserved P-loop in red, ENVG-motif in yellow and NKID-motif in dark green. Residues that are involved in GSP binding are shown as sticks, the Mg ion as green sphere. (B) The large domain of CD1219 after alignment to HypB with the aligned GSP molecule from HypB. The P-loop is shown and the conserved aspartate residue Asp153 are shown in red. (C) Alignment of CD1219 (blue) and HypB (green). P-loops are shown in red and conserved aspartate residues of CD1219 and HypB are highlighted in blue and green, respectively.

As discussed above, the location of the conserved secondary structure motifs that are characteristic for P-loop GTPases are different in the large CD1219 domain compared to traditional GTPases. Extensive literature search revealed that this permutation of conserved secondary structure motifs can also be found in other proteins (Anand *et al*., 2006). This group of proteins is called circularly permuted P-loop GTPases (cpGTPases) and includes amongst others YjeQ from *Escherichia coli* (Daigle *et al*., 2002), YloQ from *Bacillus subtilis* (Levdikov *et al*., 2004), YlqF from *Thermotoga maritima* (Kim *et al*. 2008) and YqeH from *Geobacillus stearothermophilus* (Sudhamsu *et al*., 2008). This circular permutation of secondary structure motifs affects GTP binding and hydrolysis and can lead to disturbance of the conformational stability of the protein. Therefore, many cpGTPases carry an additional N-terminal domain to stabilize the protein and the GTP binding region (Anand *et al*., 2006). This additional N-terminal domain is often involved in RNA binding. It is hence possible that RNA binding regulates GTP binding or the other way around. Furthermore, many cpGTPases contain a further stabilizing domain at the C-terminus. cpGTPases can hence be divided into different protein domains. A central GTPase domain and adjacent terminal stabilizing domains (Levdikov *et al*., 2004).

Phylogenetic analysis revealed a close evolutionary relationship between cpGTPases and bacterial ribosome-binding GTPases (Anand *et al*., 2006). The division into individual domains varies among individual cpGTPases, therefore no general architecture and domain composition of cpGTPases can be stated. The cpGTPases YjeQ and YloQ from *Bacillus subtilis* carry e.g. an N-terminal RNA-binding domain with a conserved OB-fold, an intermediate GTPase domain and a C-terminal Zinc-finger domain. In contrast to these cpGTPases, YlqF of *Bacillus subtilis* consists only of the central GTPase domain and an additional helical C-terminal domain but lacks an N-terminal stabilizing domain.

According to their three-dimensional structure and topology, these selected cpGTPases can be grouped into the TRAFAC class of GTPases (Leipe *et al*., 2002). They contain typical sequence motifs required for GTP binding, i.e. the P-loop/Walker A motif GxxxxGKS/T (G1), the Walker B motif DxxG (G3) and the N/TKxD motif (G4). Regular GTPases include the motifs in the G1-G3-G4 order in their primary amino acid sequence (Fig. 35). In cpGTPases, these motifs are circularly permuted and are arranged in the order G4-(G5)-G1-(G2)-G3. Although they have a permuted order of secondary structure elements compared to conventional GTPases, the three-dimensional structure of cpGTPases and conventional GTPases is highly similar. A sequence-based structural alignment of cpGTPases is shown in Fig. 37 and conserved GTPase motifs are highlighted in the respective colors analogous to the alignment of traditional P-loop GTPases in Fig. 35.

cpGTPases are important proteins for the respective bacteria, as they are essential for viability or growth of the bacterium, see YlqF and YqeH from *Bacillus subtilis* (Morimoto *et al*., 2002). However, the exact molecular function of most cpGTPases has not been discovered, yet.

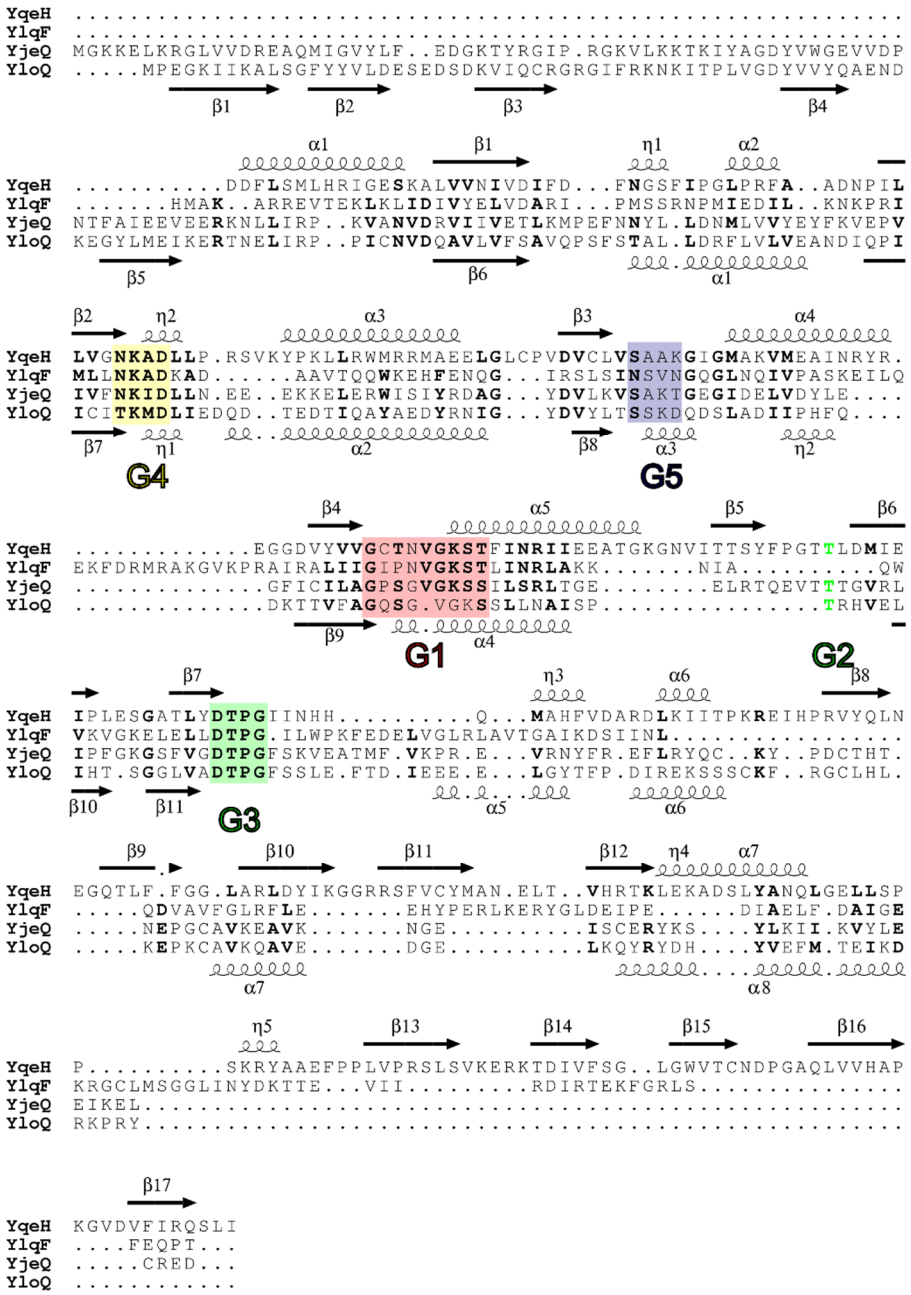

Figure 37: Structure-based sequence alignment of selected cpGTPases with highlighted motifs and residues involved in GTP binding. The alignment of YqeH from *Geobacillus stearothermophilus* (PDB-ID 3EC1, Sudhamsu *et al.*, 2008), YqlF from *Bacillus subtilis* (PDB-ID 1PUJ, Kim *et al.*, 2008), YjeQ from *Aquifex aeolicus* (PDB-ID 2YV5, Wang *et al.*, 2008, unpublished) and YloQ from *Bacillus subtilis* (PDB-ID 1T9H, Levidkov *et al.*, 2004) is shown with bold letters indicating conserved residues. Motifs involved in GTP binding are colored, with G1 (P-loop, GxxxxGKSS) in red, G2 (T) in green, G3 (DxxG) in green, G4 (N/TKxD) in yellow and G5 (SAK/L) in purple. The secondary structure elements of YqeH are shown above the alignment, these of YloQ under the alignment with β-strands as arrows and α-helices as spirals. The secondary structure elements are numbered consecutively.

Concerning their domain architecture, CD1219 from *Clostridium difficile* and the cpGTPase YloQ from *Bacillus subtilis* (PDB-ID 1T9H, Levidkov *et al.*, 2004) are highly similar. Both proteins contain an N-terminal OB-fold domain, followed by a short linker, a GTPase-like domain and a helical C-terminal domain. For YloQ it was shown that the central nucleotide-binding domain binds GTP (Cladiere *et al.*, 2006). Due to homology to GTPases, it can be assumed that also CD1219 binds GTP, but its exact molecular function and GTP-binding ability

has yet to be confirmed. The C-terminal domain of YloQ was identified as a Zn-finger domain that coordinates a Zn^{2+} ion (Levdikov *et al.*, 2004). In the case of CD1219, no Zn^{2+} ion is bound to this helical C-terminal domain. The function of the C-terminal extension in CD1219 could be a stabilizing effect on the rest of the protein that is essential due to the permutation of the GTPase motifs, similar to other cpGTPases. A comparison of the three-dimensional structure and topology of CD1219 and YloQ is shown in Fig. 38.

A detailed comparison of the topology of CD1219 and YloQ shows some striking similarities, but also differences between both proteins. The OB-fold of both proteins is highly similar. Both N-terminal domains (colored in pink in Fig. 38) consist of a β-barrel that is composed of five antiparallel β-strands. This N-terminal domain of CD1219 involves two additional short α-helices that are not present in YloQ. The OB-fold domain is linked to the central GTPase domain by a short linker (colored in yellow in Fig. 38) that consists of two short antiparallel β-strands in both proteins and an additional short α-helix in CD1219. The central GTPase domain (colored in blue in Fig. 38) of YloQ is composed of 6 β-strands (5 parallel and one antiparallel strand), flanked by α-helices on each side. CD1219, on the other hand, contains a central 8-stranded parallel β-sheet in the core domain. Due to the presence of one antiparallel β-strand in its GTPase domain, YloQ is classified as a TRAFAC GTPase and CD1219, on the other side, shares homology with SIMIBI GTPases. The C-terminal domain of YloQ (colored in cyan in Fig. 38) consists of four α-helices that bind a Zn-ion. The C-terminal domain of CD1219 involves two additional short β-strands inserted between the helices. The C-terminally bound Zn^{2+} ion of YloQ and the conserved Cys and His residues present in Zn finger domains are not present in CD1219. This indicates that the C-terminal domain of CD1219 does not possess Zn-binding affinity as compared to YloQ, but rather has a stabilizing effect on the rest of the protein.

Due to several structural and sequence similarities between CD1219 and cpGTPases, it can be assumed that CD1219 is a cpGTPase. The major difference between CD1219 and cpGTPases is the topology of the central GTPase-like domain of CD1219. Most cpGTPases belong to the TRAFAC class of GTPases, whereas CD1219 has to be classified into the SIMIBI class of GTPases due to the lack of one antiparallel β-strand that is conserved among TRAFAC GTPases.

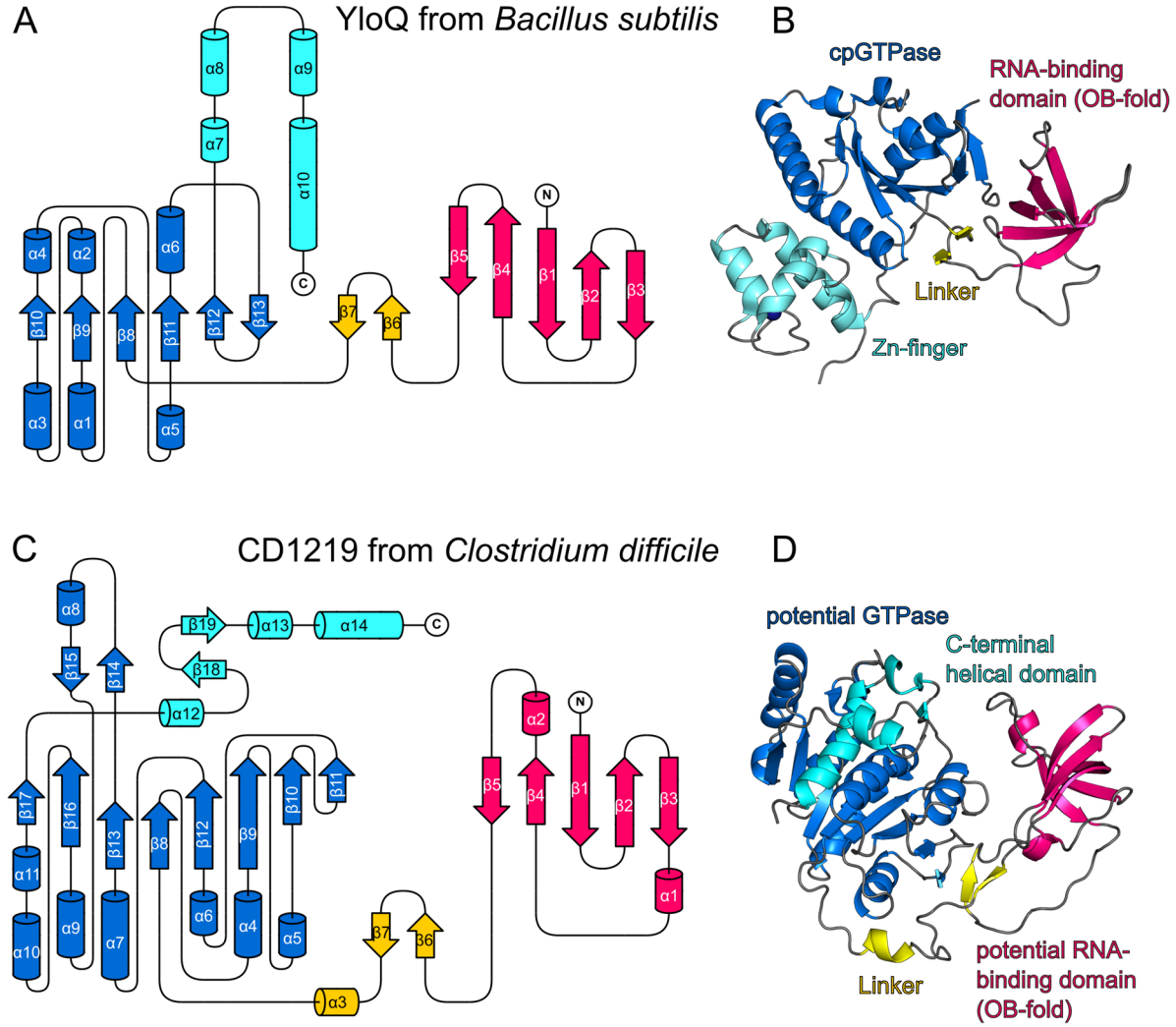

Figure 38: Comparison of the three-dimensional structure and topology of YloQ from *Bacillus subtilis* (PDB-ID 1T9H, Levdikov *et al.*, 2004) (A and B) and CD1219 from *Clostridium difficile* (C and D). The topology diagrams are shown in A and C. β-strands are illustrated as arrows, α-helices as cylinders. Topology diagrams and crystal structures are colored accordingly. The crystal structures of both proteins are shown in cartoon representation in B and D. The C-terminal OB-fold domain is colored in pink, the linker domain in yellow, the GTPase domain in dark blue and the C-terminal helical domain in cyan.

The comparison of the homologous cpGTPases YloQ from *Bacillus subtilis* and YqeH from *Bacillus anthracis* illustrates the conformational changes that take place upon GTP binding. YqeH was crystallized with a bound GTP analogon (PDB-ID 3EC1, Sudhamsu *et al.*, 2008) and YloQ in the absence of GTP (PDB-ID 1T9H, Levidkov *et al.*, 2004). An alignment of both proteins shows that conformational changes, especially in the GTP binding region, occur upon GTP binding. It is therefore possible that also CD1219 undergoes conformational changes upon nucleotide binding. The secondary structure motifs of CD1219 that are possibly involved in nucleotide binding are characterized by high B-factors (Fig. 39B). This usually indicates a high flexibility and motility of the respective region and could enable potential conformational changes in these regions upon nucleotide binding.

In the crystal structure of CD1219, a phosphate molecule is bound to each chain (Fig. 39A). The phosphate molecule is coordinated by hydrogen bonds involving the backbone residues

of the P-loop motif (T220 and G221), as well as side chains of residues of the ERHQG motif (R259 and H260) and H265. The coordination by several hydrogen bonds indicates a tight binding of the phosphate molecule. In conventional GTPases, the P-loop motif is involved in coordinating the triphosphate moiety of GTP similar to the phosphate molecule in CD1219.

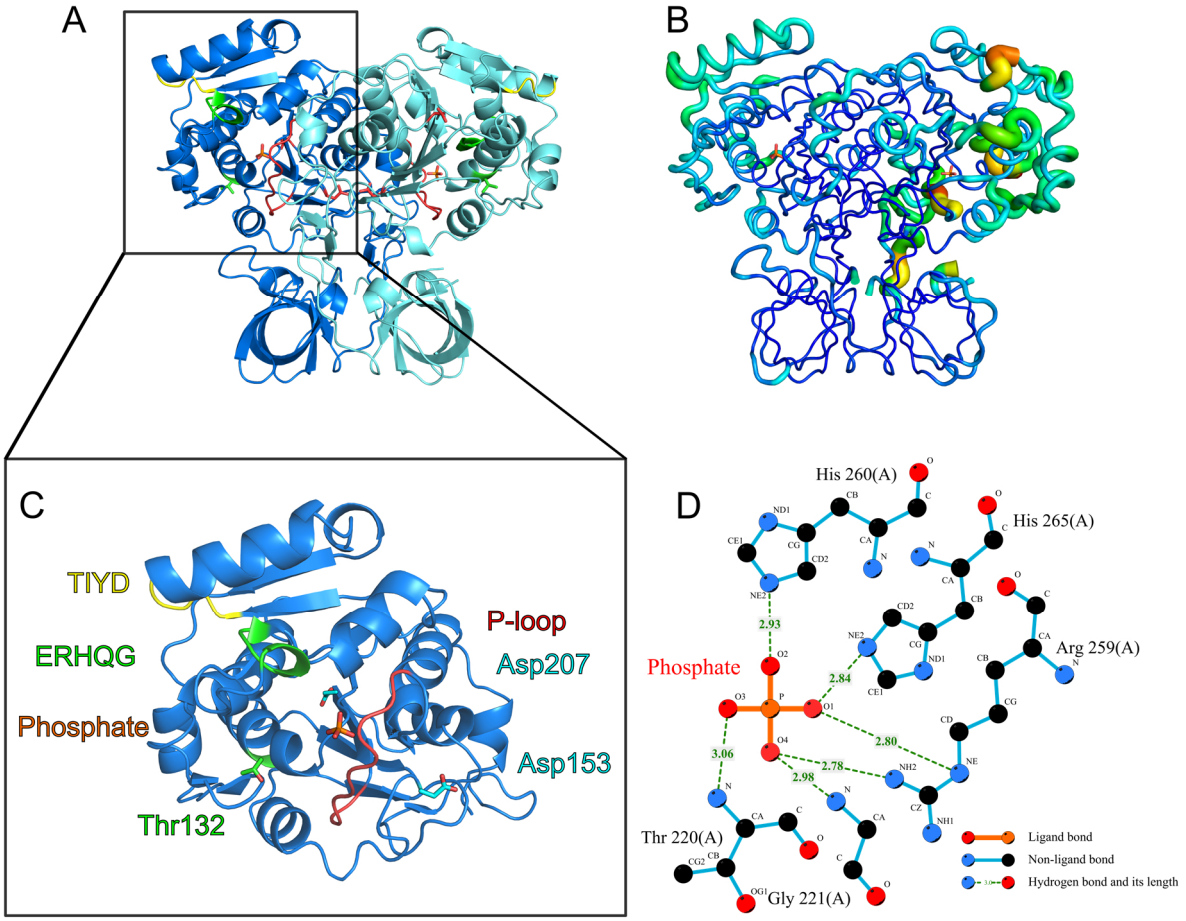

Figure 39: Conserved residues and secondary structure motifs involved in potential GTP binding in dimeric CD1219 (A) and the potential binding region in detail (C). CD1219 is shown in cartoon representation with chain A in dark blue and chain B in light blue. The TIYD motif is colored in yellow, the ERHQG motif in green, the potential P-loop in red, the bound phosphate molecule in orange, the conserved residue Thr132 in green and Asp153 and Asp207 in cyan. The structure is colored according to B-factors in (B) with red and thickened ribbon indicating high B-factors and blue highlighting low B-factors. (D) shows the Ligplot representation of the bound phosphate molecule in chain A of the crystal structure of CD1219 with the phosphate molecule colored in orange and amino acid residues in blue with carbon atoms in black, nitrogen atoms in blue and oxygen atoms in red. The length of the hydrogen bonds (colored in dashed green lines) coordinating the bound phosphate molecule is specified in Å.

This indicates that a GTP molecule could bind to CD1219 with the triphosphate part coordinated similar to the bound phosphate molecule. Analysis of the potential binding region with KVFinder (Oliveira *et al.*, 2014) reveals a binding cavity present in the crystal structure of CD1219 with a volume of ~420 Å^3 that is large enough for binding GTP (Fig. 40A). A GTP molecule was docked into this cavity using AutoDockVina (Trott & Olson, 2010). The best docking hits were analyzed manually with Pymol and the most reasonable orientation of GTP docked into the CD1219 cavity is shown in Fig. 40. Similar to the bound phosphate molecule in the CD1219 crystal structure, the triphosphate moiety of GTP is

coordinated by hydrogen bonds involving several backbone residues and side chains of the P-loop and the ERHQG motif of CD1219. The nucleotide base of the docked GTP molecule is only coordinated by hydrogen bonds involving the side chain of Thr227 and the backbone chain of Gly225. This indicates that the docked GTP molecule is tightly bound via its triphosphate moiety, but only loosely coordinated via its nucleotide base. As docking only reflects a theoretical conformation of the respective ligand and does not consider potential conformational changes of the ligand or the macromolecule, potential binding of GTP to CD1219 has to be confirmed by determination of a CD1219 crystal structure in complex with GTP or a GTP analogue.

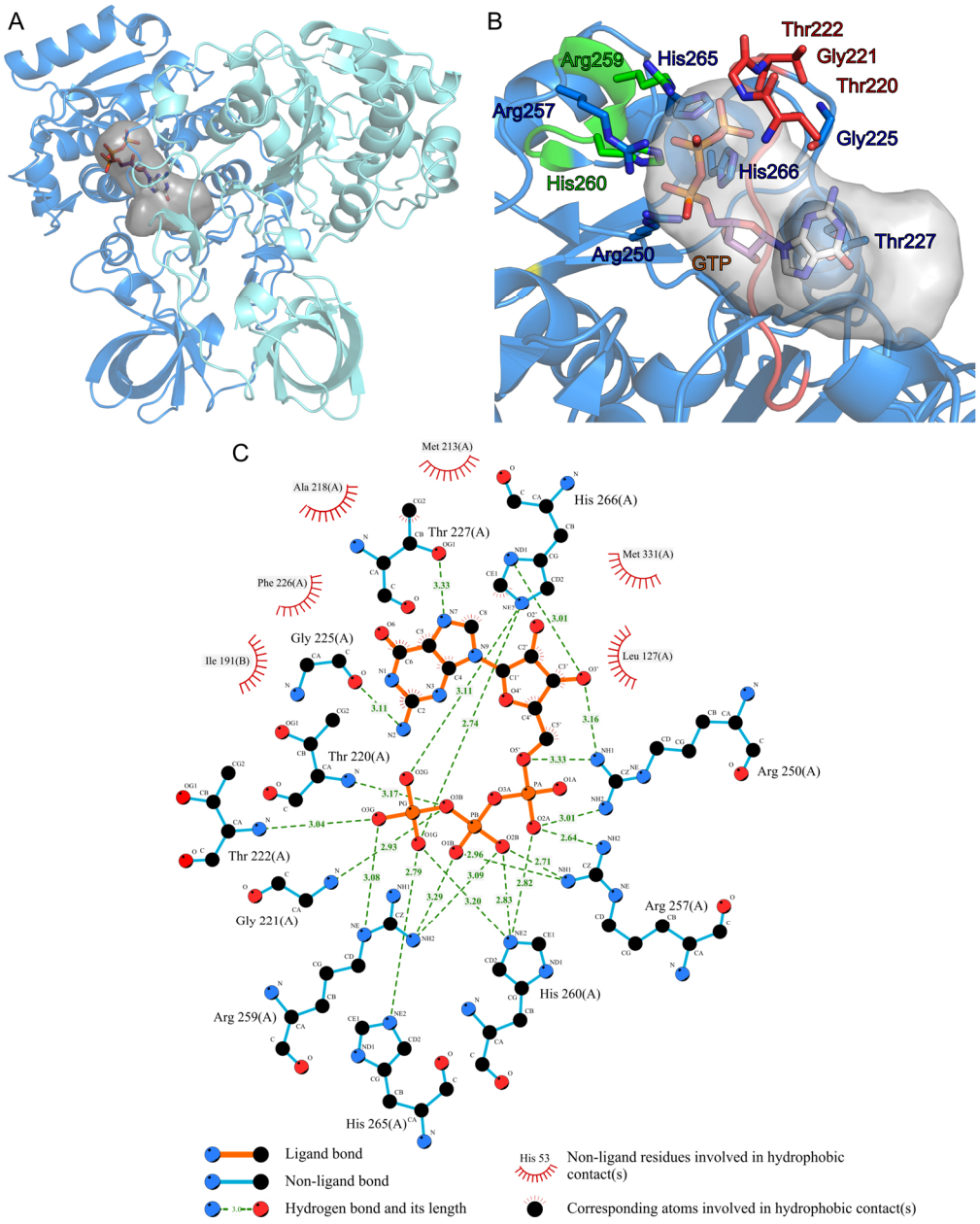

Figure 40: Potential binding of GTP to CD1219. (A) shows the binding cavity of CD1219 that was determined with KVFinder, (B) shows a detailed view of the docked GTP molecule and surrounding residues coordinating the GTP molecule. (C) shows a Ligplot illustration of the coordination of GTP by CD1219 residues. The color scheme was selected according to Fig. 39.

The conservation of the crystal structure of dimeric CD1219 and especially the region of the potential GTP-binding cavity was analyzed using ConSurf (Ashkenazy *et al.*, 2016). The conservation score was calculated by comparing the CD1219 sequence with 150 homologous sequences. The structure and amino acid sequence of CD1219 was colored according to the calculated conservation score with 1 (cyan) indicating non-conserved residues and 9 (dark red) highlighting conserved residues (Fig. 41). The residues of the linker region and the region surrounding the potential ligand binding cavity of CD1219 are highly conserved. This shows that the residues surrounding the potential ligand binding region, especially these of the P-loop, are conserved among homologous proteins of CD1219. This indicates that these proteins probably use a similar mechanism of ligand binding.

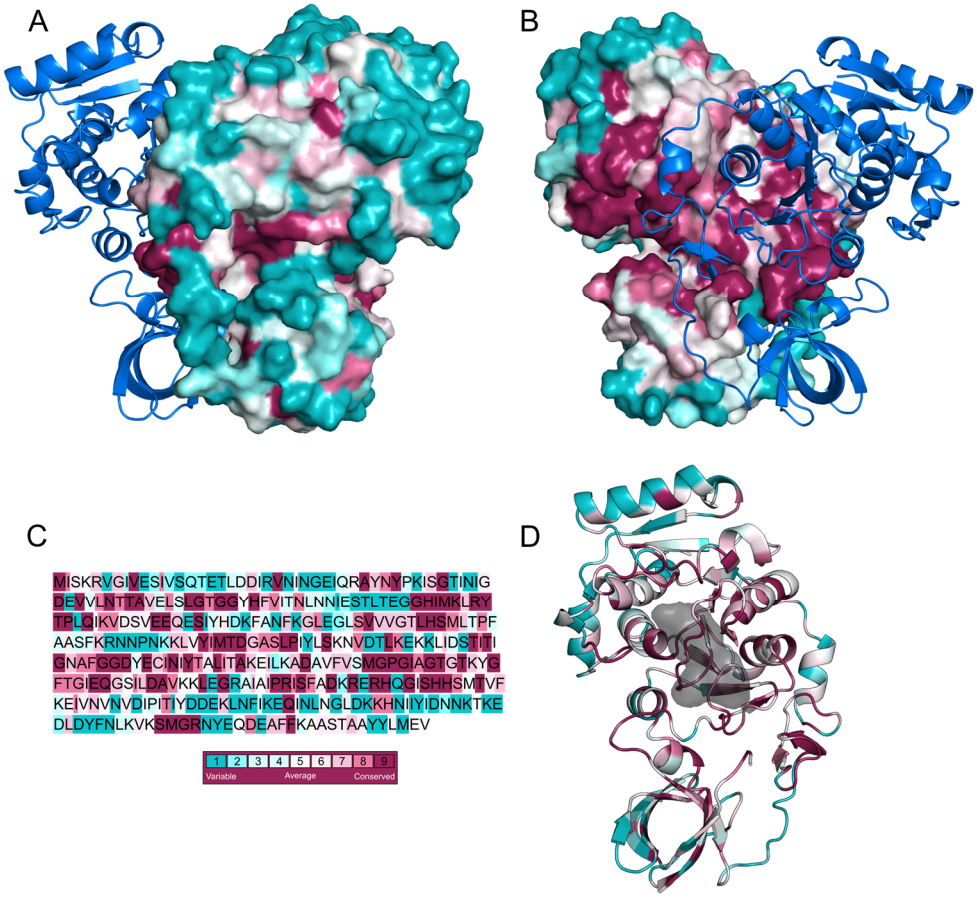

Figure 41: Conservation of CD1219 residues. (A) and (B) The surface of CD1219 is colored according to its conservation score of the respective residues. The color code is shown in (C) together with the amino acid sequence of CD1219 colored according to the score. (D) shows the potential GTP-binding cavity and the CD1219 structure in cartoon representation colored according to conservation score.

7.2.8. The genomic environment of CD1219

Analysis of the genomic environment of a gene can reveal clues about the function of its encoded protein, as proteins involved in one metabolic pathway are often encoded by genes arranged in the same gene operon (Huynen *et al.*, 2000).

DOOR (Dam *et al.*, 2007) prediction and PhyreBug analysis of the genetic environment of CD1219 revealed that the gene *CD630_12190* is arranged in a gene operon encoding several other proteins (Fig. 42). The first gene organized in this operon encodes a putative thiamine diphosphokinase (*CD630_12150*, E2.7.6.2) with ATP-binding and kinase activity. This enzyme is involved in the central thiamine metabolism and catalyzes the conversion of thiamine and ATP to AMP and thiamine diphosphate. The neighboring gene *CD630_12170* encodes a putative glucosyltransferase (E2.4.1.117), which catalyzes the conversion of UDP-glucose and dolichylphosphate to UDP and dolichyl-β-D-glucosylphosphate. *CD630_12180* also encodes a putative glucosyltransferase, but the exact molecular function of this protein is not annotated. The gene *CD630_12200* (*nudF*) encodes a putative hydrolase, involved in the reaction of ADP-D-ribose and water to AMP and D-ribose-5-phosphate (E3.6.1.13). The other genes encoded in this operon, *CD630_12160* and *CD630_12190*, are uncharacterized proteins. It is hence obvious that all genes that are encoded in this operon together with CD1219 are involved in central molecular processes, e.g. thiamine metabolism. This, together with the fact that CD1219 is essential for *C. difficile* growth *in vitro*, indicates that also CD1219 could be involved in this central metabolic pathway.

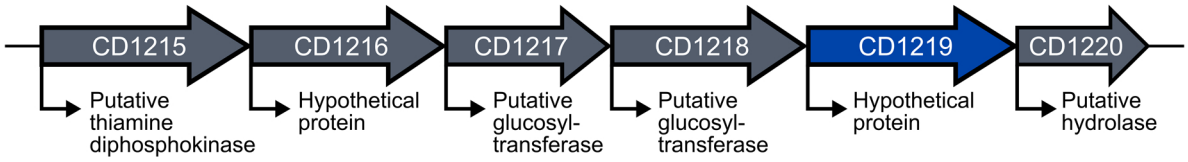

Figure 42: Operon structure of *CD630_12190* (for simplification abbreviated with CD1219). Genes are shown as arrows with arrow length illustrating gene length. The line under the respective genes summarizes the annotated function of the encoded proteins.

A STRING (Szklarczyk *et al.*, 2017) analysis of the interaction of the surrounding genes of *CD1219* illustrates that probably also genes of the neighboring operon interact with *CD1219* (Fig. 43). *CD630_12210* encodes a protein of unknown function, whereas *CD630_12220* (*xerD1*) encodes a tyrosine recombinase and *CD630_12230* (*deoB*) a phosphopentomutase. These proteins are also involved in central carbon metabolic pathways, further indicating a role of CD1219 in the central metabolism of *Clostridium difficile*.

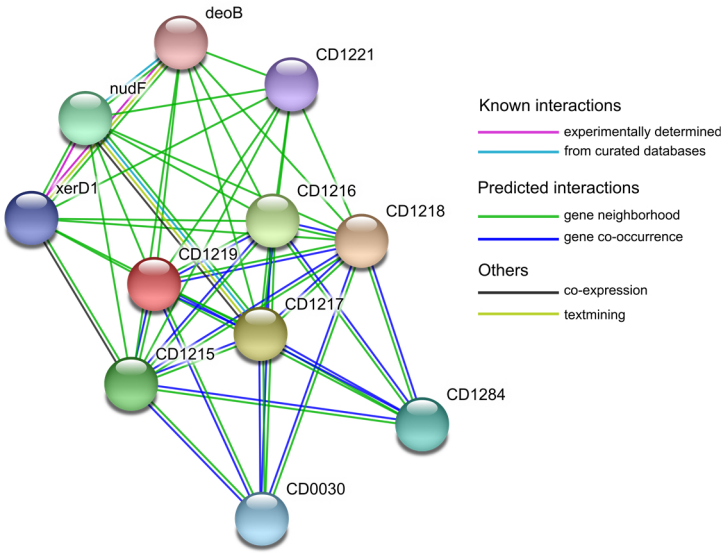

Figure 43: Interaction of CD1219 and neighboring genes. Individual genes are illustrated as nods. Interacting genes are connected by a colored line with the respective color indicating the type of interaction. The scheme was generated by STRING database.

7.3. The uncharacterized protein CD1823 of *Clostridium difficile*

The 29 kDa protein CD1823 is a structurally and functionally uncharacterized protein of *Clostridium difficile 630*. Dembek *et al*. (2015) showed that its homologue of Ribotype 20291, CD1717, is essential for sporulation of *C. difficile in vitro*. The aim of this work was to structurally characterize CD1823 and thereby derive ideas about its molecular function.

7.3.1. Structure determination of CD1823

As secondary structure prediction of CD1823 by PSIPRED (Buchan *et al*., 2013) revealed that the protein does not contain any unstructured flexible termini, full-length DNA of codon-optimized CD1823 (LifeTechnologies, Darmstadt, Germany) was cloned in various expression vectors (p10$, pET-Sumo) by SLIC cloning. The fusion protein His_6-T7-Lysozyme-CD1823 (expression plasmid p10$-CD1823) was expressed in *E. coli* BL21-CodonPlus(DE3)-RIL in TB medium. Protein expression was induced by addition of 1 mM IPTG and expression was performed at 20 °C overnight. The recombinant CD1823 construct carries an N-terminal His_6-tag, followed by T7-Lysozyme and a cleavage site for S3C-protease. His_6-T7-Lysozyme-CD1823 was purified by Ni-IMAC, tag cleavage was performed by S3C protease overnight and was followed by a reverse Ni-IMAC. A final size exclusion chromatography of CD1823 was performed to isolate pure CD1823 and to remove potential aggregates. An additional analytical size exclusion chromatography was performed to analyze the molecular weight of CD1823 by comparison to standard proteins, but the oligomeric state of CD1823 could not be determined by analytical size exclusion chromatography. The

chromatogram of the analytical size exclusion chromatography and the corresponding SDS-PAGE for analysis of the final protein purity are depicted in Fig. 44.

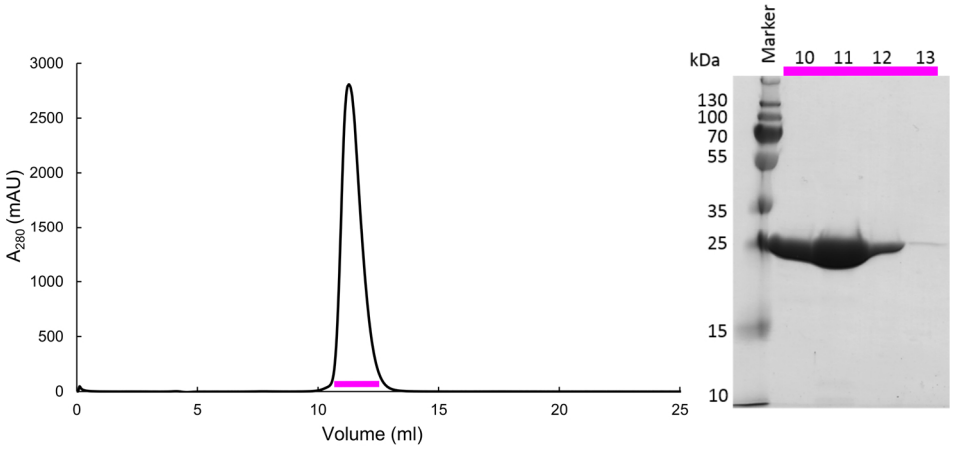

Figure 44: Chromatogram of an analytical size exclusion chromatography (S75 10/300 HR) of pure CD1823 (left) and the corresponding SDS-PAGE (right). The protein fractions that were pooled after size exclusion chromatography are marked with a pink bar. A protein standard was loaded in the first lane of the SDS-PAGE to estimate the size of the purified protein. The theoretical molecular weight of CD1823 is 29.2 kDa.

Pure CD1823 was used for crystallization experiments in concentrations of 20, 10 and 5 mg/ml using commercially available crystallization screens in sitting-drop vapour-diffusion set-ups. Crystal growth was observed in several conditions already after one day. Some crystals were fished, cryoprotected with 2,3-butanediol and immediately flash-cooled in liquid nitrogen until data were collected at beamline P11 at DESY (Deutsches Elektronen-Synchrotron, Hamburg; Burkhardt *et al.*, 2016). Exemplary CD1823 crystals are shown in Fig. 45A and the quality of the electron density of the final CD1823 structure is illustrated in Fig. 45B.

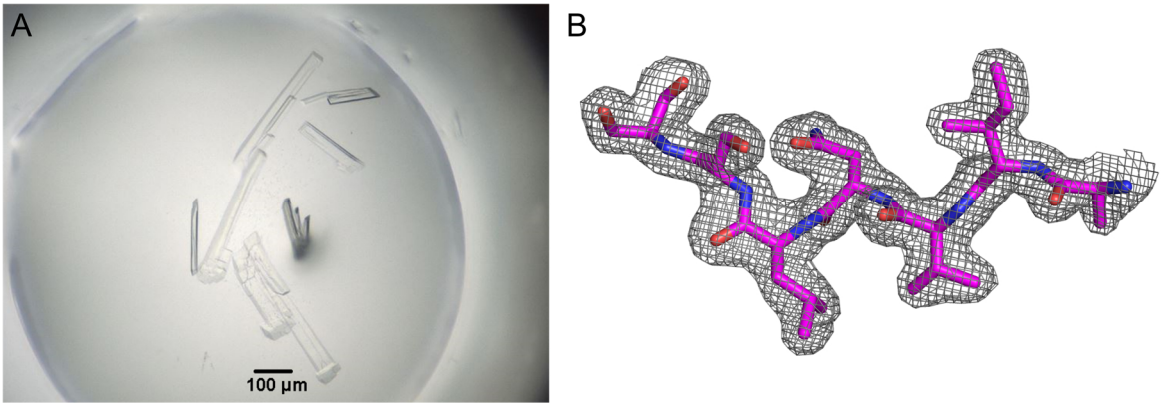

Figure 45: (A) Exemplary CD1823 crystals that were used for data collection in the crystallization condition 25% (w/v) PEG 3350, 0.1 M Tris pH 8.5, 0.2 M ammonium acetate, 5 mg/ml CD1823. (B) Exemplary $2F_o$-F_c-electron density of CD1823 at sigma contour level 1.2.

At the start of this thesis, the structure of CD1823 could not be predicted by structure prediction servers due to the lack of homologous protein structures. During the course of this thesis, the crystal structure of YaaA from *E. coli* (PDB-Code 5CAJ) was deposited in the PDB, which served as a model for MR to solve the structure of CD1823. The sequence identity of YaaA and CD1823 is 36% and MR was performed using *Phenix Phaser* (McCoy *et al.*, 2007). *Coot* (Emsley *et al.*, 2010) was used for manual model building and *phenix.refine* (Afonine *et al.*, 2012) for refinement of the structure. Data collection and refinement statistics are summarized in table 24 and table 25. CD1823 crystallized with two monomers in the asymmetric unit, indicated by a Matthews coefficient of 2.45 and a solvent content of 49.75% (calculated using the CCP4 suite). An exemplary region of $2F_0F_c$-electron density at sigma level 1.2 is shown in Fig. 45B.

Table 24: Data collection statistics for CD1823

Dataset	**CD1823_native**
Wavelength (Å) / beamline	1.033 / DESY, P11
Resolution range (Å)	49.45-1.90 (1.97-1.90)
Space group	P 1 2_1 1
Unit cell parameters (Å)	57.79 75.64 65.54
(°)	90 94 90
Mosaicity (°)	0.32
Total No. of measured reflections	296502 (19021)
Unique reflections	44441 (2861)
Multiplicity	6.7 (6.6)
Mean I/σ(I)	13.4 (1.5)
Completeness (%)	100.0 (99.9)
R_{meas} (%)	9.8 (133.4)
$R_{p.i.m.}$ (%)	3.8 (51.3)
CC1/2	99.9 (90.3)

Table 25: Refinement statistics for CD1823

Dataset	**CD1823_native**
Resolution Range (Å)	49.45-1.90 (1.97-1.90)
R_{work} (%)	19.9 (48.7)
R_{free} (%)	24.7 (51.8)
No. of non-H atoms	
Protein	4036
Ion	-
Ligand	2 (Cl^-)
Water	503
R.m.s. deviation	
Bonds (Å)	0.004
Angles (°)	0.72
Average B factors (Å^2)	
Protein	34
Ion	-
Ligand	12
Water	40
Ramachandran plot	
Favored regions (%)	98
Outliers (%)	0
MolProbity score	1.02

7.3.2. SAXS experiments of CD1823

The oligomeric state of CD1823 in solution could not be determined by analytical size exclusion chromatography, as the elution volume could not be clearly assigned to monomeric or dimeric CD1823. To determine the oligomeric state of CD1823 in solution, additional SAXS experiments were performed.

The SAXS experiments for CD1823 (in 20 mM Tris/HCl pH 8.0, 100 mM NaCl) were performed at BM29 ESRF (Grenoble, France; Pernot *et al.*, 2013) with different protein concentrations. Best data were measured for 2.5 mg/ml CD1823. The experimental scattering data were compared with theoretically calculated curves for monomeric and dimeric CD1823, while a crystallographic dimer was selected for fitting the data (Fig. 46). The *X*-value of 1.8 indicates a good fit of the experimental data with the theoretical curve for monomeric CD1823. As the fit of the experimental data and theoretical data for dimeric CD1823 yielded a higher *X*-value of 94.6, SAXS analysis of CD1823 revealed that CD1823 is present as a monomer in solution. The SAXS envelope was calculated for monomeric CD1823 and the crystal structure of monomeric CD1823 was fitted into the SAXS envelope using Chimera (Fig. 46C).

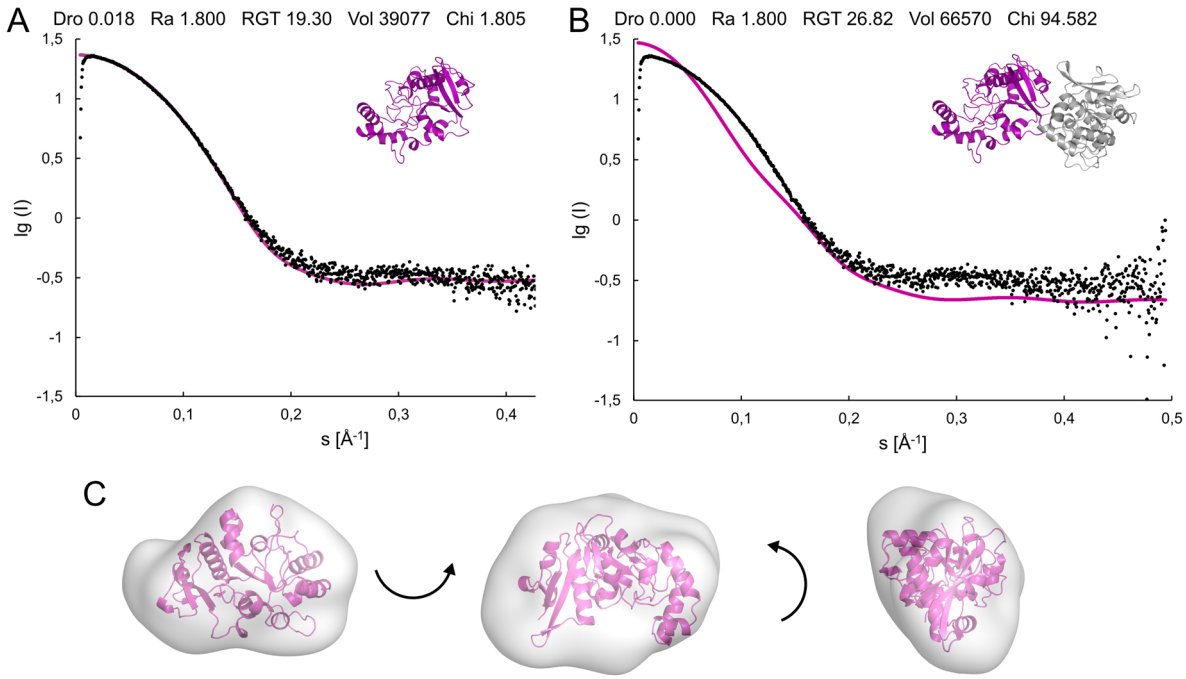

Figure 46: SAXS analysis of CD1823. The fit of the experimental data (black dots) and theoretical SAXS curves (magenta curves) for monomeric (A) and dimeric CD1823 (B) is shown in the top panel. The crystal structure of monomeric and dimeric CD1823 is shown as cartoon representation with chain A in magenta and chain B in grey. (C) The fit of the calculated SAXS envelope for monomeric CD1823 and the crystal structure of monomeric CD1823 in magenta cartoon representation is shown in the bottom panel of the figure in different orientations.

7.3.3. Crystal structure of CD1823

CD1823 crystallized with two monomers in the asymmetric unit. As the CD1823 dimer is only a crystallographic and no functional dimer, the structural analysis of the protein will be performed with monomeric CD1823 in the following chapters.

Monomeric CD1823 consists of one protein domain with a central β-sheet composed of five parallel β-strands that are flanked by additional five antiparallel strands, while strand β7 is significantly longer compared to the other strands. Many α-helices are located around the central β-strands on each side. CD1823 does not contain any conserved and characteristic secondary structure motifs that are conserved among traditional protein families. The crystal structure of CD1823 and the corresponding two-dimensional topology diagram are depicted in Fig. 47 with β-strands in magenta and α-helices in grey.

Analysis of the electrostatic surface of monomeric CD1823 shows a large positively charged patch present on two neighboring sides of CD1823, which is illustrated in blue color in Fig. 47. The other sides of CD1823 show an even distribution of positive and negative charge without any special peculiarities.

A small cavity can be found in CD1823 with a volume of ~230 Å^3. The surface of the region surrounding this cavity is positively charged, indicating potential binding of a negatively charged ligand in this cavity. In the crystal structure that was determined in this thesis, this

cavity is occupied by several water molecules and a chloride ion, but is large enough to accommodate a small ligand instead of water molecules.

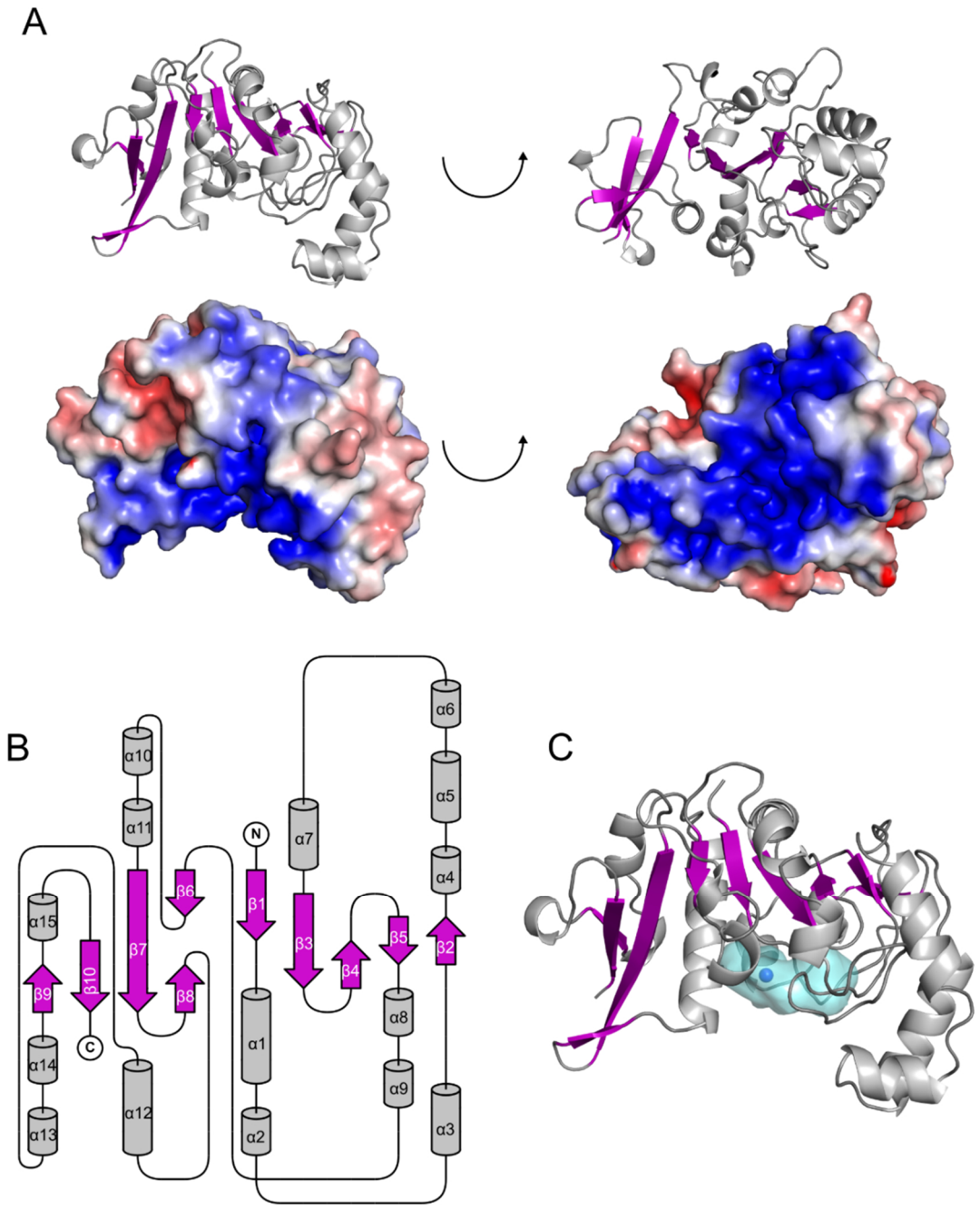

Figure 47: (A) Crystal structure and electrostatic surface charge distribution of monomeric CD1823 of *Clostridium difficile*. Central β-strands are colored in magenta, surrounding α-helices in grey. The top panel illustrates the crystal structure of CD1823 and the bottom panel the electrostatic surface in the respective orientation of the structure in the panel above. The right panel is a 90° turned representation of the left orientation. Blue patches of the surface indicate positive charged, red patches indicate negative charged parts of the surface. (B) Topology diagram of monomeric CD1823. β-strands are simplified as arrows and α-helices as cylinders. The secondary structure elements are numbered consecutively. (C) Potential ligand binding cavity of CD1823 shown as cyan surface with bound Cl- ion (shown as blue sphere).

The conservation of CD1823 was calculated between CD1823 and 150 homologous sequences using ConSurf (Ashkenazy *et al.*, 2016). The structure and amino acid sequence of CD1823 was colored according to conservation score from 1 to 9 with 1 showing variable and 9 the highly conserved regions (Fig. 48). Variable regions are colored in cyan, conserved regions are highlighted in dark red. Especially the region surrounding the potential ligand binding cavity of CD1823 including the bound chloride ion is surrounded by highly conserved residues. This indicates that this region is conserved among homologous proteins and could indeed depict a potential ligand binding region. Based on the homologous proteins that were selected for calculation of the conservation score, no hypothesis about the potential function of CD1823 could be derived, as the function of these proteins has also not been annotated so far.

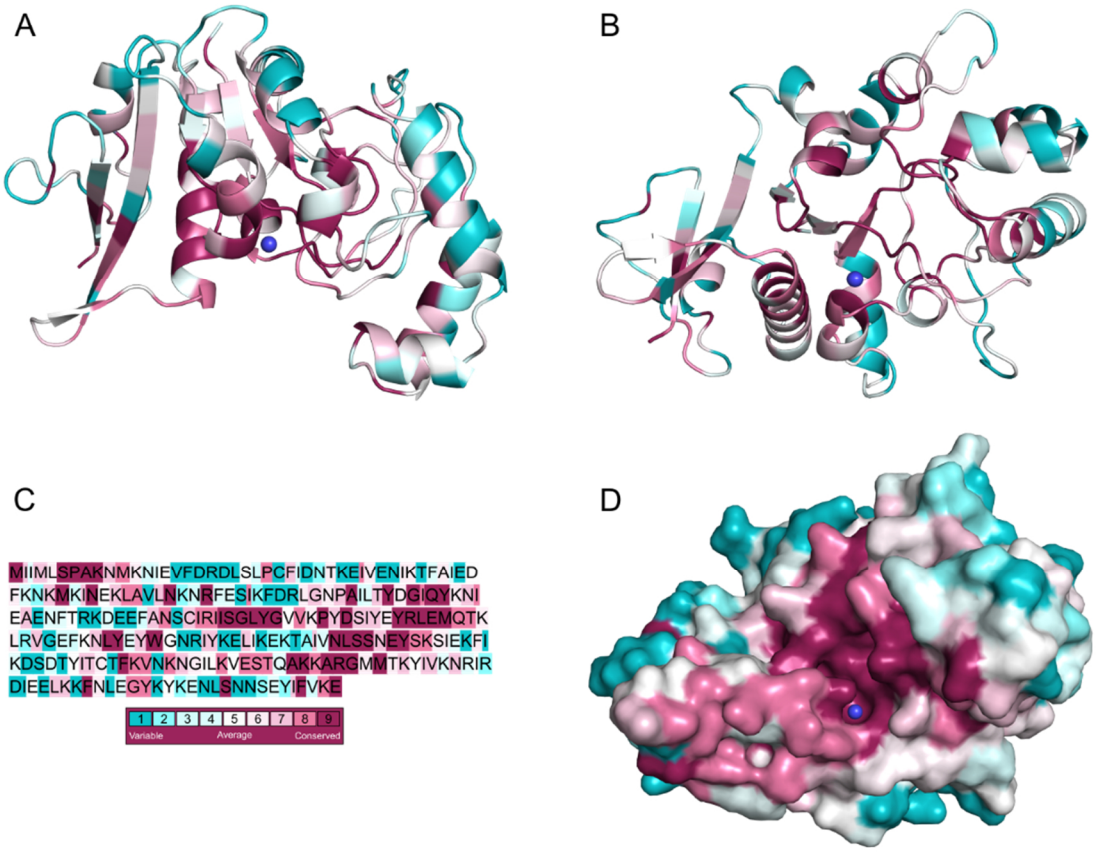

Figure 48: Conservation of CD1823. The structure of CD1823 is colored according to the conservation color code in (C) with 1 (cyan) showing variable and 9 (dark red) showing conserved residues. The structure of CD1823 is shown in cartoon representation in different orientations in (A) and (B) and as surface representation in (D). The bound chloride ion is shown as blue sphere.

7.3.4. Comparison of CD1823 and homologous proteins

The structure of CD1823 was determined by MR using the structure of YaaA as search model. YaaA from *E. coli* and CD1823 from *C. difficile* share a sequence identity of 36% and an identical protein fold. Both structures can be aligned with an r.m.s.d. of 1.23 Å over 248 residues. A structure-based sequence alignment is shown in Fig. 49A and a structural alignment of both structures is illustrated in Fig. 49B with CD1823 in magenta and YaaA in light cyan. Both proteins show almost identical three-dimensional structures, with only minor differences. YaaA contains two short β-strands (β2 and β4) that are not present in CD1823. Latter, on the other hand, possesses a short C-terminal α-helix (η5) that can't be found in YaaA.

The electrostatic surface distribution of YaaA is similar to that of CD1823 (Fig. 47 and Fig. 50). Both proteins contain a highly negative charged, continuous patch on one side of the protein structure, while the rest of the surface shows an even distribution of positive and negative charge. Similar to CD1823, YaaA has a bound chloride ion in the same position as in CD1823, which is surrounded by a highly positive charged surface.

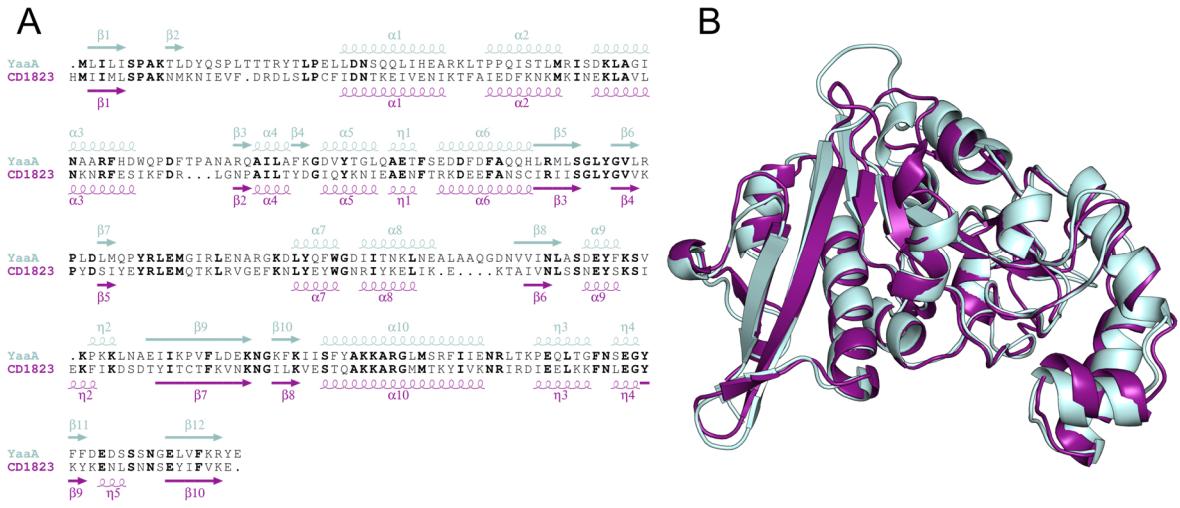

Figure 49: (A) Structure-based sequence alignment of CD1823 from *Clostridium difficile* and YaaA from *Escherichia coli* (PDB-ID 5CAJ, unpublished). Conserved residues are shown in bold and the secondary structure elements of YaaA are shown above the alignment, CD1823 elements are depicted below the alignment. α-helices are depicted as spirals, β-strands as arrows. Secondary structure elements are numbered consecutively. (B) Structural alignment of the crystal structures of CD1823 (purple) and YaaA (light cyan).

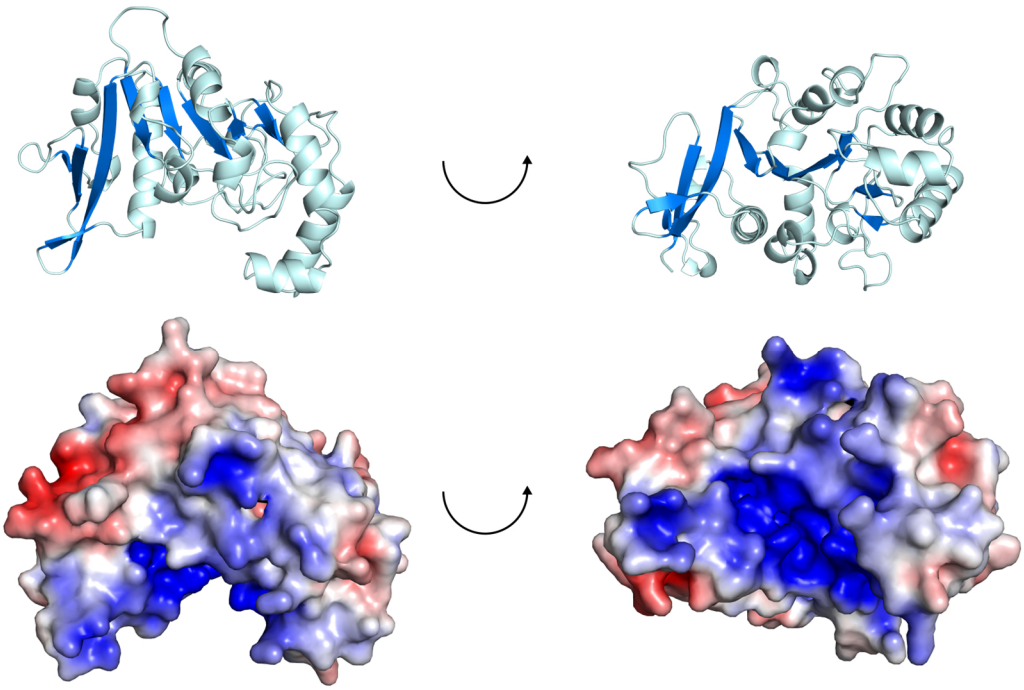

Figure 50: Electrostatic surface of YaaA. The crystal structure of YaaA is shown in cartoon representation in two orientations and as electrostatic surface in the same orientations. The second orientation (right panel) is turned by 270° around the x-axis with respect to the first one (left panel).

Due to the high structural similarity between YaaA and CD1823, hypotheses about the molecular function of CD1823 could be derived based on the function of YaaA. YaaA from *E. coli* is annotated as a protein involved in peroxide stress response (Liu *et al.*, 2011). Upon peroxide stress response, reactive oxygen species (ROS) can be generated intracellularly under aerobic conditions or extracellularly upon inflammation processes in the gut. They can enter cells due to membrane penetration and can cause severe damage by disturbing several central metabolic biomolecules. H_2O_2 is typically produced after reaction of oxygen with metals or thiols. Under normal conditions, iron and hydrogen peroxide undergo the so-called Fenton reaction (Fenton, 1894) and ROS are produced, which can cause DNA lesions and inactivation of iron enzymes or iron-sulfur enzymes (Liu *et al.*, 2011).

Upon peroxide stress response, YaaA is expressed under control of the OxyR regulon in H_2O_2 presence. The OxyR regulon is activated when an activated thiolate residue of the OxyR transcription factor is oxidized by H_2O_2 to a sulfenic acid. This activated transcription factor then directly triggers expression of peroxide stress response proteins, e.g. YaaA (Lee *et al.*, 2004; Zheng *et al.*, 1998). YaaA, together with Dps and Fur, lowers the intracellular levels of unincorporated iron and thereby helps to attenuate the Fenton reaction and to prevent DNA damage caused by reactive oxygen species (ROS), e.g. H_2O_2 (Liu *et al.*, 2011). The exact molecular mechanism of YaaA and its potential role in iron metabolism is still unknown. Several potential mechanisms how YaaA might lower unincorporated iron levels inside the cell have been postulated. On one hand, it could slow down the import of iron into the cell and could on

the other hand also prevent release of iron from H_2O_2-damaged enzymes. Furthermore, it is possible that YaaA could accelerate the transfer of free iron to target proteins or increase its efflux out of the cell (Liu *et al.*, 2011).

In anaerobic bacteria, such as *C. difficile*, peroxide stress response is initiated in presence of oxygen. *C. difficile* expresses several proteins upon peroxide stress response, among these are manganese superoxide dismutase (CD1631), several putative manganese catalases (CD1567, CD0598, and CD2401) and a putative thioredoxin-dependent thiol peroxidase (CD1822) (Janoir *et al.*, 2013).

The high similarity between YaaA and CD1823 indicates a similar function of both proteins. As YaaA is involved in peroxide stress response, it can be assumed that also CD1823 is involved in this process. Because the exact molecular function of YaaA and its role during peroxide stress response are currently unknown, no hypotheses about the molecular function of CD1823 can be derived from its *E. coli* homologue.

YaaA and CD1823 belong to the UPF0246 protein family, which has not been functionally characterized so far. Computational analysis using InterPro (Finn *et al.*, 2017) showed that over 8700 other proteins from various organisms (archaea, aerobic and anaerobic bacteria, eukaryota) belong to this protein family. Sequence comparison of some of these protein family members showed several conserved amino acid patches (Fig. 51). However, as the exact molecular function of proteins of the UPF0246 family is not known, no ideas about the function of CD1823 can be derived from this homology.

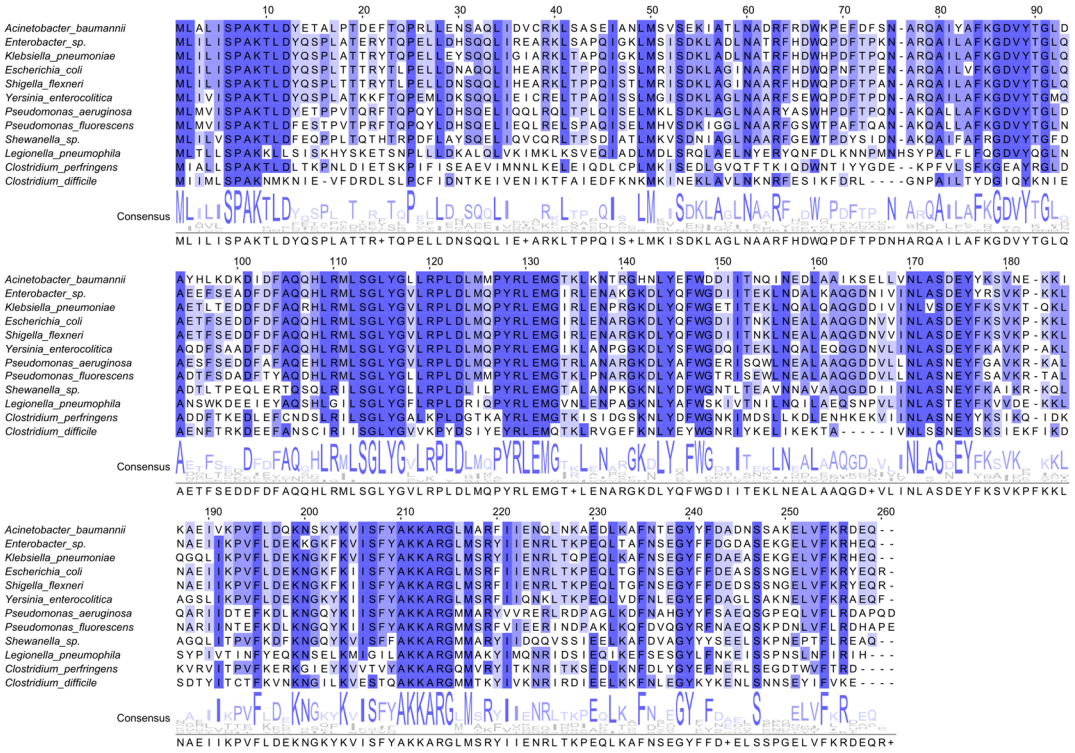

Figure 51: Sequence alignment of CD1823 from *Clostridium difficile* and other proteins of the UPF0246 protein family. Amino acids that are conserved in > 80% of the aligned proteins are colored in dark blue, a conservation in > 60% of the selected proteins is illustrated by mid blue and a conservation in > 40% in light blue.

The structure of CD1823 was submitted to online structural alignment servers such as DALI (Holm, *et al.*, 2016) and PDBeFold (Krissinel & Henrick, 2004) to search for structurally homologous proteins. A search with PDBeFold revealed the structural homologue YaaA from *E. coli* (PDB-Code 5CAJ) that was already discussed above. DALI search also revealed YaaA of *E. coli* as the best hit, but also showed some proteins with lower sequential and structural similarity to CD1823 with a dramatic decrease in the alignment Z-score and r.m.s.d. compared to YaaA. The second-best hit was ManxA from *Thermoanaerobacter tengcongensis* (PDB-ID 3LFH, Fu & Su, 2010, unpublished), part of a phosphotransferase system, which could be aligned with 42% CD1823 residues with an r.m.s.d. of 3.1 Å and showed an overall Z-score of only 4.8. A small part of the central core of CD1823 could be aligned with ManxA, but despite the similar topology of this small region, no hypothesis about the function of CD1823 could be derived based on the structural comparison with ManxA.

7.3.5. Genetic environment of CD1823

To get ideas about the potential function of CD1823, its genomic context was analyzed. The gene *CD630_18230* is located at position 2,112,409 in the genome of *Clostridium difficile* strain 630. Gene operon analysis by DOOR (Mao *et al.*, 2009) predicted that the gene *CD630_18230* is organized in an operon together with the neighboring gene *CD630_18220* (*bcp*), which encodes a thioredoxin-dependent thiol peroxidase (E1.11.1.15). This protein is actively involved in peroxide stress response, as it is expressed upon peroxide stress, shows peroxidase activity and catalyzes the reduction of H_2O_2 (Janoir *et al.*, 2013).
Another gene, *CD630_18240*, which is not organized in the same operon as CD1823, encodes a p-type calcium transport ATPase. It flanks the predicted *CD1823* operon upstream. Downstream to the *CD630_18230* operon is another operon, including the genes *CD630_18200* and *CD630_18210*. These genes encode an adenine deaminase (*ade*) and a transposase-like protein B (*tlpB*), respectively. The operon organization and genetic environment of *CD630_18230* is illustrated in Fig. 52.

Figure 52: Genomic context of *CD630_18230*. According to simplification, gene names are abbreviated with *CD18230* instead of *CD630_18230*. Arrows of the same color indicate genes belonging to one operon.

Analysis of the genomic context and operon organization of homologous genes of *CD1823* in other bacterial strains, e.g. *Pseudomonas aeruginosa* and *Legionella pneumophila*, showed no conservation of the *CD1823* operon structure. Thus, no potential function of CD1823 could be derived based on comparison of its operon structure with the operon structure of homologous genes of related bacterial strains.

The genetic context of CD1823 was additionally analyzed using the STRING database (Szklarczyk *et al.*, 2017). The result is shown in Fig. 53. Due to gene neighborhood, it is assumed that *CD1823* interacts with gene *CD1824* (a cations-transporting ATPase), *CD1822* (*bcp*, a thiol peroxidase), *CD1820* (*ade_2*, an adenine deaminase), *CD0049* (*proS1*, a prolyl tRNA-synthetase), *CD0050* (*proS2*, a prolyl tRNA-synthetase), *CD2275* (*topB2*, a DNA topoisomerase III), *CD1701* (*recQ*, an ATP-dependent DNA helicase) and *CD1700* (*ribD*, a riboflavin biosynthesis protein).

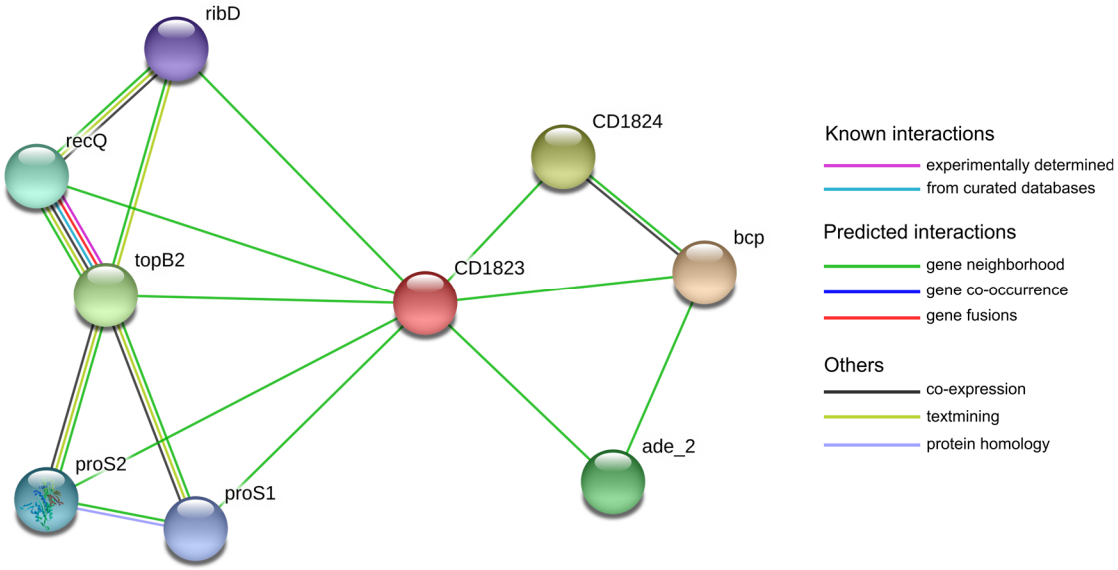

Figure 53: Analysis of the genetic context of CD1823 using the STRING database. The genes are depicted as colored nods and the line between two genes illustrates an interaction of the respective genes. The line color illustrates the type of interaction.

The gene neighborhood of CD1823 and the putative thiol peroxidase CD1822, as well as the structural similarity between CD1823 and YaaA from *E. coli*, a peroxide resistance protein, indicates that CD1823 could also be involved in peroxide stress response. However, the exact molecular function of CD1823 needs still to be determined.

7.4. Other essential proteins of *Clostridium difficile*

Besides CD1219 and CD1823, three other essential, uncharacterized proteins of *Clostridium difficile* were selected for structural and functional analysis (table 18). While the structures of CD1219 and CD1823 could be determined by X-ray crystallography, the structural analysis of the other three proteins was more complicated. All proteins were cloned in expression vectors and test expression was performed analogous to CD1219 and CD1823. All cloned constructs are listed in table 38 in the supplement.

The fusion constructs of CD1067 with an N-terminal His_6-MBP, His_6-T7-Lysozyme or His_6-Sumo-tag, respectively, could be solubly expressed in *E. coli*, but tag cleavage was not successful, as the recombinant protein precipitated after tag removal. An optimization of the

construct boundaries and attached tags is essential to find a suitable construct for protein expression and purification. A TDSA might help to find a buffer composition that is better suited for protein purification and tag cleavage of CD1067. Another option could be to perform crystallization experiments without cleaving off the affinity tag.

The uncharacterized essential protein CD2589 of *C. difficile* was successfully purified in this thesis and pure protein was used for initial crystallization screens. Several crystals were subjected to diffraction experiments, but diffraction (resolution > 20 Å) was not sufficient for data collection. Optimization of the crystallization conditions using random screens and grid screens resulted in an improvement of the resolution to 10 Å. Crystals of SeMet-labeled CD2589 diffracted to 6 Å, but this resolution was still not sufficient for structure determination. A CD spectrum of native CD2589 was collected and is shown in Fig. 84 in the supplement. This spectrum reveals that the protein is properly folded in the buffer that was used for protein purification. To further improve diffraction quality of the crystals, a surface entropy reduction (SER) mutant of CD2589 was cloned. Therefore, three flexible residues (Glu121, Lys122 and Glu123), which are predicted to be located on the surface of the protein, were mutated to rigid alanine residues. Expression and purification of the SER mutant was performed analogous to native CD2589. However, the SER-mutant of CD2589 eluted in a single peak with small shoulders during size exclusion chromatography, indicating that the SER-mutant is not stable, but rather present in several oligomeric states. The SER-mutant of CD2589 was hence not used for crystallization setups. An optimization of buffer composition for purification of the SER mutant has to be performed to yield a stable protein sample that can be used for crystallization experiments. Further optimization of the crystal conditions, together with limited proteolysis of native CD2589 and SER-CD2589 might also help to improve the diffraction quality of CD2589 crystals.

The essential protein CD2752 was cloned in various expression vectors and all fusion proteins could be soluble expressed in *E. coli*. Purification of the protein yielded a stable protein sample of CD2752, but several impurities were still present in the sample after a 4-step purification. An optimization of the purification protocol and buffer is hence necessary to obtain pure protein that can be used for crystallization experiments.

Overall, the structure of two essential proteins of *Clostridium difficile*, CD1219 and CD1823, could be determined in this thesis and could pave the way for further experiments for a functional analysis of these proteins. The structure of the other three selected proteins of this project could not be determined due to purification and diffraction problems. The general workflow and success of the structural and functional analysis is summarized in Fig. 54.

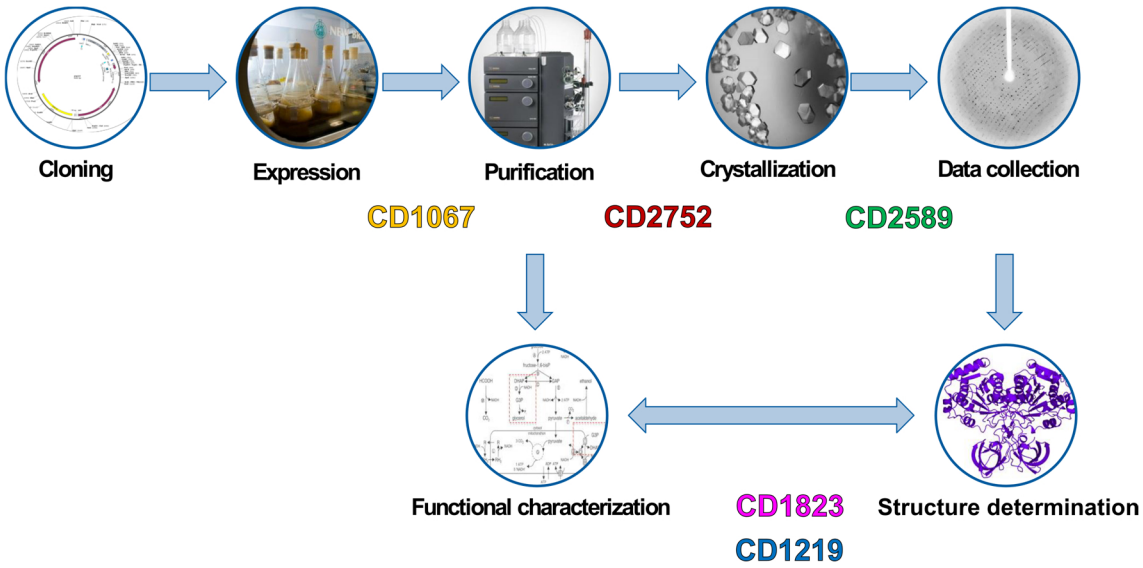

Figure 54: Overview of the workflow and progress of the structural and functional analysis of essential uncharacterized proteins of *Clostridium difficile*.

8. Diffocins

The so-called diffocin proteins of *Clostridium difficile* are high-molecular weight, bacteriophage tail-like bacteriocins. Due to their killing activity towards related strains in their environment, they are treated as novel potential therapeutic agents for CDI. They are structurally related to other phage-tail like particles.

8.1. Phage-tail like particles

8.1.1. Contractile bacteriophage tails

The architecture of bacteriophages is rather simple, as they consist in principle just of two parts: a phage capsid and a contractile tail. The capsid carries the genomic phage DNA that is transferred into the host cell during infection. The phage tail is involved in penetration of the host cell membrane and pore formation but it plays no active role in the DNA transfer. Upon host cell attachment, the bacteriophage tail undergoes large conformational changes that result in contraction of the tail envelope. Upon contraction, the inner tail part is pushed through the host cell membrane. The mechanism of action of the bacteriophage tail can hence be compared to a syringe (Leiman *et al.*, 2004).
According to their architecture, bacteriophage tails can be grouped into three different classes (Ackermann 2003):

- *Myoviridae*: long contractile tail
- *Siphoviridae*: long non-contractile tail
- *Podoviridae*: short non-contractile tail

Bacteriophage tails consist of several structural elements (Fig. 55). The baseplate carries several tail fibers, also called receptor binding proteins (RBPs) that are involved in receptor binding and host cell attachment (Arisaka *et al.*, 2016). A tube complex is attached to the baseplate that is surrounded by the contractile sheath, followed by the terminator complex, which connects the phage tail and capsid (Leiman & Shneider, 2012). The first step of the host-cell infection involves receptor recognition and binding, which is supported by digestion of external polysaccharides present on the host cell surface due to the enzymatic activity of the RBPs. The peptidoglycan layer is then destroyed by a glycosidase present in the bacteriophage tail (Leiman *et al.* 2007, Walter *et al.* 2008). Upon contact to the host cell membrane, the baseplate undergoes a conformational change, which results in sheath contraction to half or even less of its original length. This pushes the inner part, the tail tube, through the host cell membrane, which creates a channel for transfer of the phage DNA (Leiman *et al.*, 2010). It could be shown by cryo-EM studies (Leiman *et al.* 2004, Kostyuchenko *et al.* 2005) that the T4 sheath can be present in an elongated and a contracted state.

The tail of bacteriophage T4 is one of the best-studied bacteriophage tails (Leiman *et al.*, 2009). They are approximately 220 Å in diameter, but variable in their length (1.000 - 4.500 Å). Their structural organization is rather simple, as they are composed of only two components: the internal rigid tube (proteins with 15 - 21 kDa) and the external contractile sheath (proteins with 40 - 80 kDa) are both composed of copies of one single protein, respectively. The genes encoding the sheath and the tube protein are usually located adjacent to each other, with the sheath gene preceding the tube gene. Their transcription is usually regulated by other genes, transcription regulators (Leiman & Shneider, 2012).
Besides bacteriophages, there are several homologous protein machineries that share a phage-tail like architecture, e.g. the pyocins from *Pseudomonas aeruginosa*, the bacterial Type VI secretion system (T6SS) or proteins encoded by the *Photorhabdus* virulence cassette. The function of these contractile tail systems differs between the individual systems, as they can be involved in transport of protein or DNA, but also in transfer of ions into the cytoplasm of the host cell.

8.1.2. Pyocins from Pseudomonas aeruginosa

Pyocins from *Pseudomonas aeruginosa* are high-molecular weight bacteriocins (Jacob 1954). According to their definition, bacteriocins are proteins or peptides that are produced by bacteria to kill other specific bacteria in their environment (Cotter *et al.*, 2013). The pyocin genes are usually transcribed upon SOS response, e.g. after UV or mitomycin treatment (Matsui *et al.*, 1993). This leads to DNA damage and activation of RecA. RecA degrades the PrtR protein, a repressor of PrtN, which can then activate transcription of the pyocin genes (Michel-Bryand & Baysse, 2002).
According to their architecture, pyocins are bacteriophage tail-like particles. The pyocin particle consists of 11 proteins that are encoded in the pyocin gene cluster, similar to bacteriophages. Based on their architecture, pyocin particles can be grouped into three different classes (Michel-Bryand & Baysse, 2002):

- R-type pyocins: non-flexible, contractile phage-tail like particles
- F-type pyocins: flexible, non-contractile phage-tail like particles
- S-type pyocins: colicin-like proteins

R-type pyocins are one of the smallest contractile tail systems. One bacterium can produce up to 200 pyocin particles that are released into the environment upon lysis of the bacterial cell. However, one pyocin particle is enough to kill a bacterium. The bacterium that produces the pyocin particles is resistant against its own pyocin. It is assumed that production of pyocins is used to ensure survival of a bacterial species under stress conditions. One bacterial cell has

to sacrifice itself and lyse to release the pyocin particles to ensure the survival of its sister cells (Lee *et al*., 2016).
Due to their structural similarity to contractile bacteriophage tails and to diffocin particles, the focus in this thesis will be on R-type pyocins; F- and S-type pyocins will not be discussed.

8.1.3. Other phage-tail like particles

Several other phage-tail like particles are known, e.g. the contractile tail of the bacterial T6SS or proteins encoded by the virulence cassette of *Photorhabdus* (Fig. 55). The bacterial T6SS is a complex macromolecular machine that is involved in pathogenesis. It consists of several proteins, among which several share a structural and sequence similarity to phage tail proteins (Leiman 2009). The contractile T6SS tail consists of the tube protein Hcp, while the sheath is composed of two proteins, VipA and VipB. The Hcp tube can be up to 1000 Å long with a diameter of 85 Å (Bönemann 2010). Similar to phages and pyocin particles, the T6SS uses its contractile tail to attack host cell membranes and to form a pore (Alteri & Mobley, 2016). In contrast to bacteriophages, this channel is used for the transfer of large proteins into the host cell cytoplasm. Other differences between T6SS and phage tails are e.g. that the T6SS baseplate consists of only three instead of five proteins, no tape measure protein is present in the T6SS and its sheath is build up by two proteins instead of only one protein. Furthermore, the bacterial T6SS has to act continuously and therefore has to be recycled (Leiman & Shneider, 2012).
Photorhabdus sp. produces proteins encoded in the so-called *Photorhabdus* virulence cassettes (PVC) that are similar to R-type pyocins. In contrast to pyocins, the PVC proteins are not classified as bacteriocins, as they do not show any antibacterial activity (Yang *et al*., 2006). Instead, they are toxic towards eukaryotic cells. The PVC proteins are encoded in a PVC gene operon that encodes tail sheath proteins and several other T6SS-related proteins, e.g. a VgrG-like protein (Hurst *et al*., 2004; Pukatzki *et al*., 2007). Similar to R-type pyocins, the PVC particle can be present in a contracted and an elongated state. It can therefore be assumed that the mechanism of pyocins and PVC proteins is similar and that also the PVC proteins drive the inner tube through the target cell membrane upon sheath contraction.

8.1.4. Diffocin proteins as potential therapeutics for *C. difficile* infections

Upon SOS-response, *Clostridium difficile* produces the so-called diffocin proteins (Gebhart *et al*., 2012). These proteins are high-molecular weight bacteriocins and their architecture is similar to tail proteins of the *Myoviridae* phages, R-type pyocins of *Pseudomonas aeruginosa* and the bacterial T6SS (Fig. 55). Due to their high selectivity for specific *Clostridium difficile* strains, the diffocin particles could be used as therapeutic agents for *C. difficile* infections, as they can also kill antibiotic-resistant strains (Sangster *et al*., 2014). The diffocin that is produced

by one specific *C. difficile* strain does not kill the same bacterial strain, but other specific strains of the same species. A strain producing diffocins is hence typically resistant against its own diffocins (Gebhart *et al*., 2012).

The target specificity of the diffocin particle is determined by its receptor binding proteins (tail fibers), which differ between the individual diffocins (Gebhart *et al*., 2012). By swapping the RBPs and engineering new diffocin proteins, one could target the killing-activity to a certain *C. difficile* strain and use the diffocin proteins as a potential therapeutic. It could already be shown that engineered diffocin proteins stay active after oral admission and after passing the gastrointestinal tract of mice. They do also not damage the gut microbiota as compared to "traditional" antibiotic treatment (Gebhart *et al*., 2015).

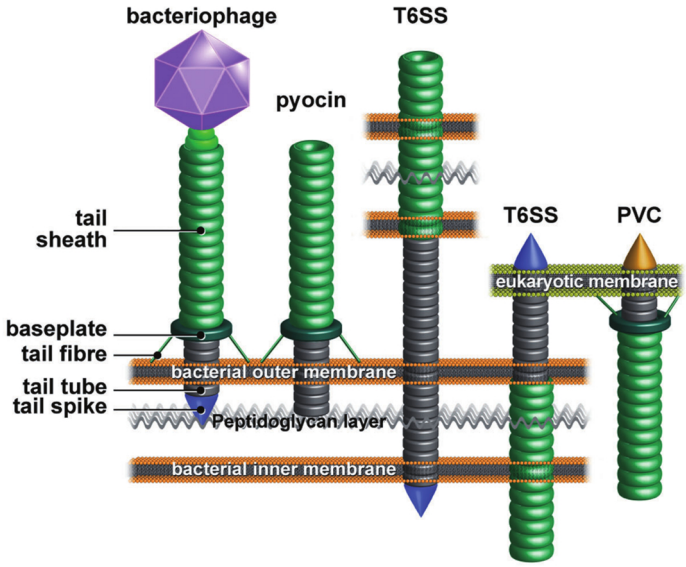

Figure 55: Schematic architecture of contractile bacteriophage tail-like systems. The figure is adapted with permission from Bönemann *et al*., 2010.

In *Clostridium difficile* CD630, the diffocin proteins are encoded in an operon that contains 25 predicted open reading frames (ORFs), including structural and regulatory proteins, as well as proteins with unknown function (CD13590-CD13760, Gebhart *et al*., 2012). The potential function of the respective proteins was derived from homology to known and well-characterized *Myoviridae* phage genes. The operon can be found in all sequenced *C. difficile* strains and the individual genes distribute a high homology among the individual strains, apart from the highly variable genes encoding the tail fibers and receptor binding proteins (CD1373 and CD1374) (Gebhart *et al*., 2012). The diffocin gene cluster is shown in Fig. 57 and the respective diffocin genes of CD630 encoded in this operon are summarized in table 26.

The diffocin tail is composed of a tube surrounded by a contractile, non-flexible sheath, which is arranged on a baseplate, followed by six tail fibers and a spike protein. Upon SOS-response, these diffocin particles are produced inside the bacterium. At a certain time point, the bacterial cells lyse and release the diffocin particles into the environment. Upon binding of the tail fibers

to unique receptors on the host cell surface, the sheath contracts and pushes the inner tube through the bacterial membrane, thereby creating a pore (Gebhart *et al.*, 2015). Due to homology to phage tail proteins, it is assumed that the baseplate changes its conformation, which causes sheath contraction and insertion of the tube through the membrane (Kostyuchenko *et al.*, 2003). The created pore is big enough for transfer of ions, but its size is still small enough that no larger molecules such as proteins or toxins can get out of the host cell. This leads to disturbance of the membrane potential and depolarization of the cytoplasmic membrane, which kills the target bacterium. The architecture of the diffocin particle in its pre- and post-contraction state and the proposed diffocin mechanism is shown in Fig. 56. As the structure of the main components of contractile systems (T6SS, T4 phage tails, pyocin particle) are conserved, it is assumed that they also share a similar mechanism.

However, the origin of bacteriophage-like tails has not been discovered, yet. As the structure of the core elements of the contractile tails is quite conserved without a clear homology at the level of the amino acid sequence, it is still unclear whether they have emerged from phages or bacteria (Leiman & Shneider, 2012).

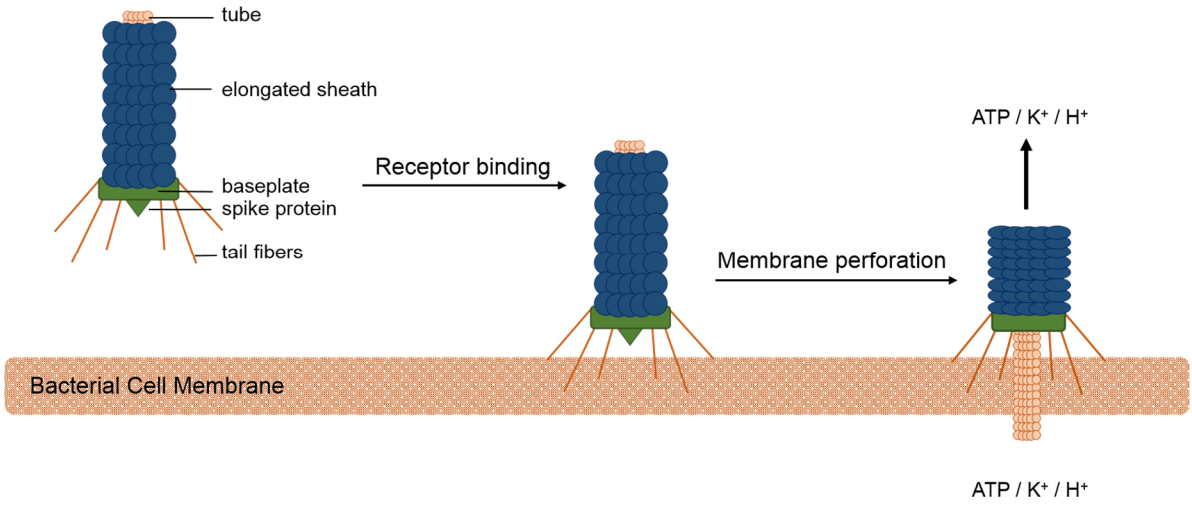

Figure 56: Overview of the schematic architecture of the diffocin particle in its pre- (left) and post-contraction state (right) and the diffocin mechanism. Upon binding to receptors on the host cell surface, the diffocin sheath contracts, which pushes the inner tube through the target membrane. A pore is formed, through which ions can be transferred out of the bacterial cell. The host cell is then killed due to disturbance of the membrane potential.

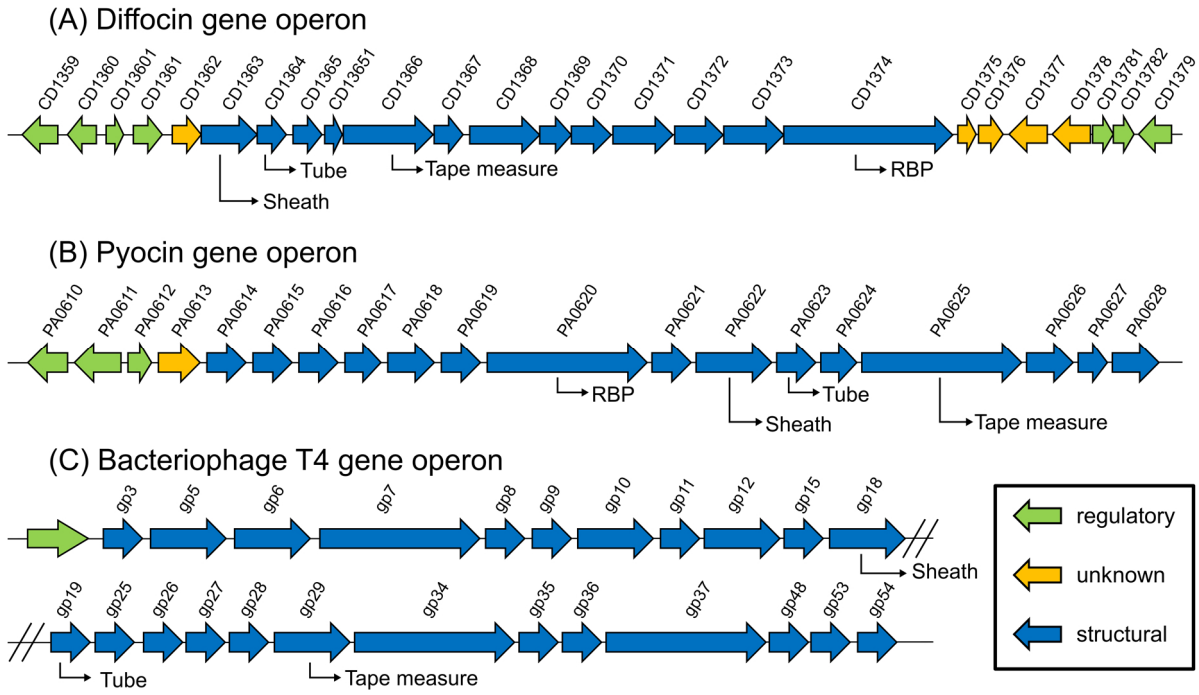

Figure 57: Gene clusters of different phage-tail like particles. The diffocin gene cluster is shown in (A), the pyocin gene cluster in (B) and the gene operon of bacteriophage T4 tail components in (C). Genes are illustrated as arrows with the gene name above the arrow. The arrow color indicates the annotated function (green for regulatory proteins, yellow for proteins with unknown function and blue for structural proteins). The gene cluster of bacteriophage T4 tail components is reduced severely to its basic components for simplification. Several regulatory and phage capsid genes were left out for clarity.

Table 26: Genes of the diffocin operon with size of the respective encoded protein and their annotated function

Gene	Size (aa)	Info
CD13590	168	Putative phage protein, regulatory
CD13600	106	Transcriptional regulator, Phage-type, regulatory
CD13601	65	Transcriptional regulator, Phage-type, regulatory
CD13610	146	Putative phage protein, regulatory
CD13620	147	Uncharacterized protein, unknown function
CD13630	354	Putative phage XkdK-like protein, structural (sheath)
CD13640	142	Putative phage XkdM-like protein, structural (core)
CD13650	148	Putative phage XkdN-like protein, structural
CD13651	55	Uncharacterized protein, structural
CD13660	817	Putative phage tail protein, structural (tape measure)
CD13670	140	Putative phage XkdP-like protein, structural
CD13680	509	Putative phage cell wall XkdQ-like hydrolase, structural (baseplate)
CD13690	108	Putative phage protein, structural
CD13700	142	Putative phage XkdS-like protein
CD13710	350	Putative baseplate assembly predicted J-like, structural (baseplate)
CD13720	232	Putative phage XkdT-like protein, structural
CD13730	328	Putative phage tail fiber protein, structural (tail fiber)

CD13740	1743	Putative beta-lactamase-inhibitor protein II, structural (receptor binding)
CD13750	39	Putative phage protein, unknown function
CD13760	86	Putative phage protein, unknown function
CD13770	151	Putative membrane protein, unknown function
CD13780	139	Transcriptional regulator, Phage-type, regulatory
CD13781	69	Transcriptional regulator, Phage-type, regulatory
CD13782	73	Uncharacterized protein, regulatory
CD13790	132	Transcriptional regulator, Phage-type, regulatory

8.2. Aim of this project

At the beginning of this project, the diffocin proteins of *Clostridium difficile* were structurally uncharacterized. Negative stain images of the diffocin particles are available (Sangster *et al*., 2014), but crystal structures of diffocin proteins have not been determined, yet. Cryo-EM structures of the homologue R-type pyocins of *Pseudomonas aeruginosa* were published in 2015 (Ge *et al*., 2015; PDB-Codes 3J9R and 3J9Q) in the pre- and post-contraction state at 3.5 and 3.9 Å, respectively. These structures illustrate the overall shape of the pyocin sheath complex and highlight the interaction of the individual monomers in the assembled particle. Due to the lack of a high-resolution crystal diffocin structure, a detailed atomic structural characterization was not possible so far.

To get insight into the structural details of the individual diffocin proteins, the structures of the non-oligomerized sheath and tube protein have been investigated by X-ray crystallography and SAXS experiments in this thesis. Comparison with homologous structures, e.g. pyocins and phage tail proteins reveals potential similarities and differences between diffocin and other phage-tail like particles and provides an opportunity to obtain insight into the assembly of the contractile apparatus of diffocins and other R-type bacteriocins as well as construction of a model of the diffocin particle in its assembled state.

8.3. Structural analysis of the diffocin sheath protein CD1363

The 39 kDa protein CD1363 from *Clostridium difficile* is a structural protein encoded in the diffocin gene operon. It is annotated as a putative phage XkdK-like protein building the sheath envelope of the diffocin particle (UniProt-Code Q18BN0), but its exact molecular function and structure were unknown at the start of this project.

8.3.1. Structure determination of CD1363

The gene encoding CD1363 (*CD630_13630*) was amplified from genomic *Clostridium difficile* CD630 DNA (provided by Prof. Dr. Ralf Gerhard, MHH, Germany) and was cloned in a

pOPIN-M expression vector by SLIC-cloning. The fusion construct His_6-MBP-S3C-CD1363 was solubly expressed in *E. coli* BL21-CodonPlus(RIL) cells by autoinduction at 25 °C for 24 h. Purification of CD1363 was achieved by affinity chromatography (MBP-tag), followed by tag cleavage by S3C-protease overnight, another step of affinity chromatography (reverse Ni(II)-IMAC) and final size exclusion chromatography. Pure protein was used for crystallization setups using commercially available screens. Crystal growth was obtained in several conditions already after one day. As no suitable model was available for structure solution by molecular replacement, SeMet-labeled CD1363 was expressed in M9 minimal medium in *E. coli* BL21-CodonPlus(RIL) cells in TB-medium at 20 °C overnight. Purification and crystallization setups were performed analogous to native CD1363. The chromatogram of the final size exclusion chromatography step of SeMet-labeled CD1363 and the corresponding SDS-PAGE illustrating the final protein purity is depicted in Fig. 58A and B. Exemplary crystals of native and SeMet-labeled CD1363 are shown in Fig. 58C and D.

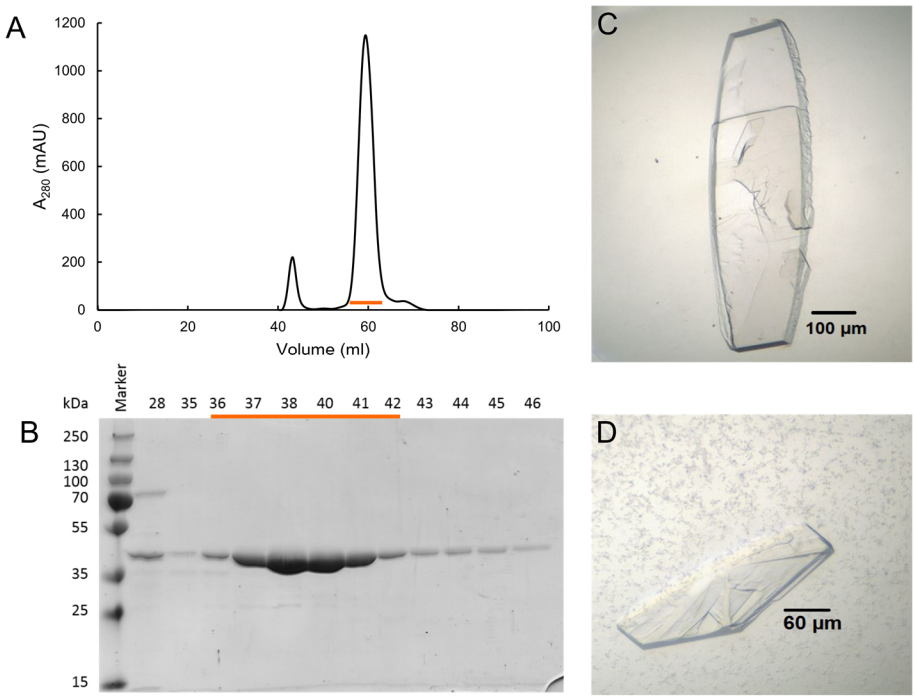

Figure 58: Chromatogram (A) and corresponding SDS-PAGE of the final size exclusion chromatography of SeMet-labeled CD1363. A protein standard is loaded in the first lane of the SDS-PAGE for comparison of the molecular weight of CD1363 (theoretical MW CD1363 = 39 kDa) with standard proteins. Protein fractions that were pooled and used for crystallization setups are marked with an orange bar. Exemplary crystals of SeMet-labeled CD1363 (C) and native CD1363 (D). Crystallization condition (C): 1.0 M Lithium chloride, 0.1 M Tris pH 8.0, 20% (w/v) PEG 6000, 20 mg/ml CD1363; Crystallization condition (D): 0.1 M BIS-TRIS pH 5.5, 25% (w/v) PEG 3350, 20 mg/ml CD1363. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

Crystals were fished directly from the crystallization screens, 2,3-Butanediol was used as cryoprotectant and data were collected at beamline PXII at SLS (PSI, Villigen, Switzerland; Fuchs *et al.*, 2014). Statistics of the dataset collection of SeMet-labeled CD1363 crystals are summarized in table 27. The crystal structure was solved using *Autosol* (Adams *et al.*, 2010)

from the *Phenix* suite, initial model building was performed using *Autobuild* (Terwilliger *et al.*, 2008), followed by iterative refinement cycles (using *Phenix.Refine* (Afonine *et al.*, 2012)) and manual model building using *Coot* (Emsley & Cowtan, 2004; Emsley *et al.*, 2010). Refinement statistics are shown in table 28.

Table 27: Dataset collection statistics for SeMet-labeled CD1363.

Dataset	**CD1363 SeMet**
Wavelength (Å) / beamline	0.97945/ SLS, PXII
Resolution range (Å)	38.36-1.90 (1.97-1.90)
Space group	C121
Unit cell parameters (Å)	75.94 47.88 112.51
(°)	90.00 92.25 90.00
Mosaicity (°)	0.23
Total No. of measured reflections	432693 (27784)
Unique reflections	32089 (3199)
Multiplicity	13.5 (13.3)
Mean I/σ(I)	21.8 (1.8)
Completeness (%)	99.8 (99.6)
R_{meas} (%)	9.2 (159.3)
$R_{p.i.m.}$ (%)	2.5 (43.0)
CC1/2 (%)	100 (70)

Table 28: Refinement statistics for CD1363.

Dataset	**CD1363 SeMet**
Resolution Range (Å)	38.36-1.90 (1.96-1.90)
R_{work} (%)	21.11
R_{free} (%)	24.57
No. of non-H atoms	
Protein	2689
Water	183
R.m.s. deviation	
Bonds (Å)	0.004
Angles (°)	0.57
Average B factors (Å^2)	
Protein	51
Water	46
Atoms	51
Ramachandran plot	
Favored regions (%)	97.4
Outliers (%)	0
MolProbity score	1.64

8.3.2. Crystal structure of CD1363

CD1363 crystallized in space group C121 with one molecule in the asymmetric unit. The solvent content is approximately 53% with a Matthews coefficient of 2.61 (calculated with CCP4_Matthews_coef). The crystal structure of monomeric CD1363 and the corresponding two-dimensional topology diagram is shown in Fig. 59.

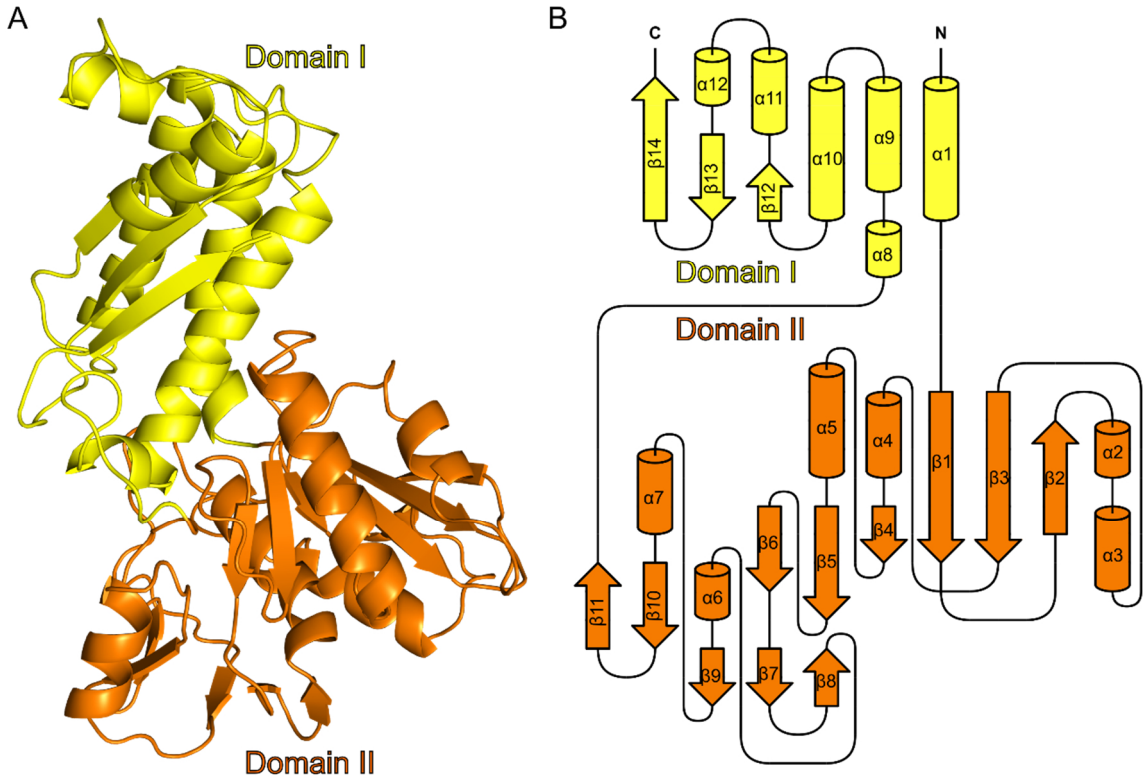

Figure 59: (A) Crystal structure of monomeric CD1363 from *Clostridium difficile* in cartoon representation. Individual domains are colored in orange and yellow, respectively. (B) Topology diagram of monomeric CD1363. The domains are colored analogous to the cartoon representation of CD1363 in (A). Secondary structure elements are numbered consecutively; α-helices are illustrated as cylinders and β-sheets as arrows. The figure was prepared using Topdraw (Bond, 2003). Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

Monomeric CD1363 is composed of two domains, domain I (residues 4-25 and 226-354) and domain II (residues 26-225), colored in yellow and orange in Fig. 59, respectively. It is important to note that domain II is an insertion into domain I. Domain I starts with a long α-helix at the N-terminus, followed by domain II, which folds back into domain I at the C-terminus. Domain I is composed of a central three-stranded β-sheet with two parallel and one antiparallel β-strand. The β-sheet is sandwiched between six α-helices. Domain II (colored in orange) contains a central six-stranded β-sheet with five parallel and one antiparallel β-strand, flanked by four α-helices. The β-sheet is followed by five additional smaller β-strands and two α-helices.

The surface of CD1363 shows an equal distribution of alternating negative and positive charged patches (Fig. 60). This allows an interaction of CD1363 monomers due to attraction of opposite surface charges and could favor the assembly of the individual monomers into the

diffocin particle. CD1363 contains several negatively charged “fingers” and small positively charged cavities on its surface that could interact with each other due to electrostatic attraction.

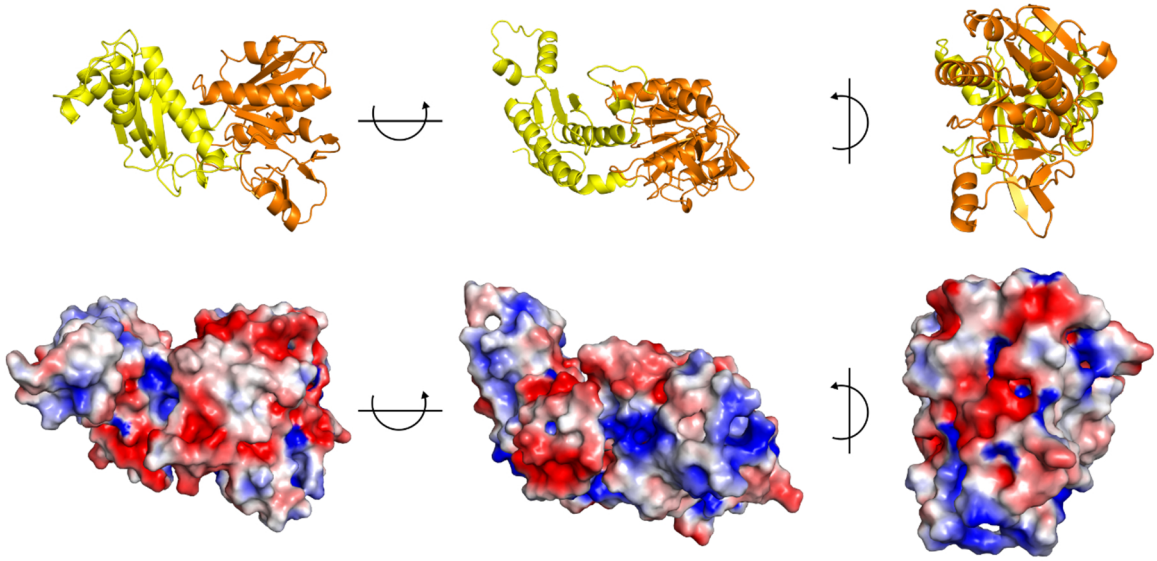

Figure 60: Electrostatic surface of the crystal structure of CD1363 from *Clostridium difficile* in different orientations. Red patches illustrate negatively charged surface regions, blue patches indicate positive charge. Each orientation is shown in cartoon representation (top panel), as well as electrostatic surface (bottom panel). Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

A large cavity flanked by several residues of both domains and with a volume of ~ 580 Å^3 is present in the interface between domain I and II (Fig. 61). The outer part of this cavity, which is directly accessible from the solvent, is surrounded by several positively charged amino acids, e.g. lysine and arginine. The inner part is mainly negatively charged due to the presence of several glutamate residues.

The conservation of CD1363 was analyzed using the ConSurf server (Ashkenazy *et al.*, 2016). The crystal structure of CD1363 was colored according to the calculated conservation score and is illustrated in Fig. 62. Cyan patches depict non-conserved residues, dark red patches indicate conserved residues. The amino acids of the central β-sheet of domain II as well as the helical parts of domain I are highly conserved. In contrast to this, the central β-sheet of domain I and the helices of domain II are highly variable in their amino acid composition.

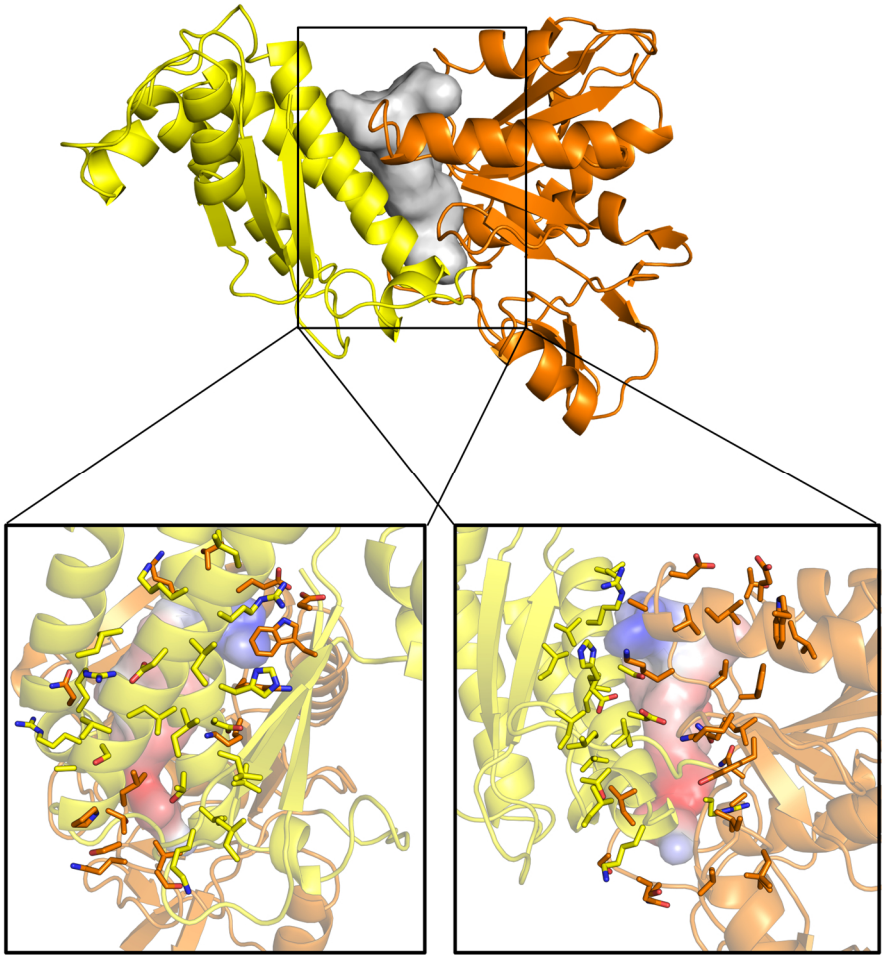

Figure 61: A cavity is present in the interface between domain I and II of CD1363 from *Clostridium difficile*. The cavity is shown as light grey surface representation in the top panel with the crystal structure of CD1363 in cartoon representation. Domain I is colored in yellow and domain II in orange. The bottom panel illustrates the electrostatic charge distribution of the cavity in two orientations with flanking residues in stick representation. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

Figure 62: Crystal structure and amino acid sequence of CD1363 colored according to the calculated conservation score. A score of 1 illustrates variable residues, a score of 9 depicts conserved amino acids. The scores were calculated by the Consurf server. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.3.3. SAXS analysis of CD1363

SAXS experiments of CD1363 were performed at BM29 ESRF (Grenoble, France; Pernot *et al.*, 2013) to analyze the oligomeric state of the diffocin sheath protein in solution. The best data were obtained for 0.63 mg/ml CD1363 in 20 mM HEPES pH 7.5, 150 mM NaCl. Experimental SAXS data are shown as black dots in Fig. 63. These experimentally measured data were fit against theoretically calculated scattering curves for monomeric and dimeric CD1363. A crystallographic dimer of CD1363 was selected for this purpose. The fit for monomeric CD1363 is shown in Fig. 63A and is described by a *X*-value of 1.8. The fit of dimeric CD1363 yielded a *X*-value of 25.7 and is shown in Fig. 63B. Due the low *X*-value of the fit for monomeric CD1363, SAXS experiments showed that CD1363 is monomeric in solution. The SAXS envelope was hence calculated for monomeric CD1363 and the crystal structure of the monomer was fit into the envelope using Crysol (Fig. 63C).

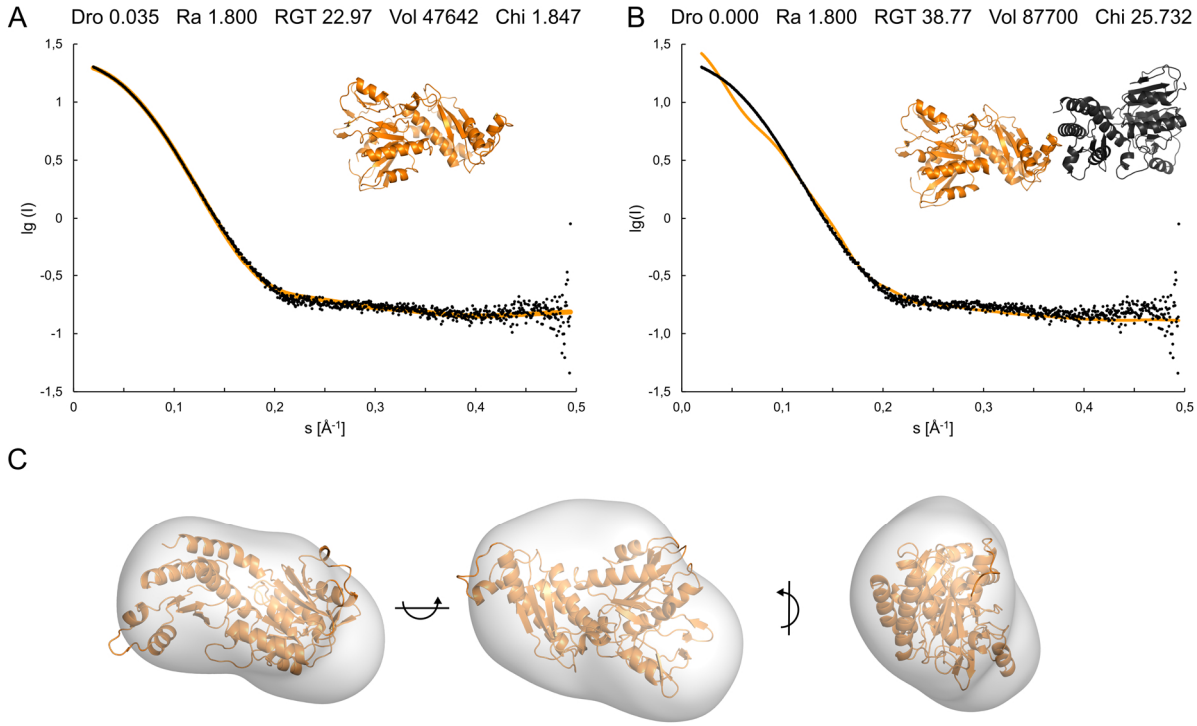

Figure 63: Fit of experimental SAXS data of CD1363 (black dots) and theoretical SAXS data (orange curves) for monomeric (A) and dimeric (B) CD1363. (C) Fit of the crystal structure of monomeric CD1363 (orange, cartoon representation) into the calculated SAXS envelope (surface representation) in different orientations (turned by 90° around x- and y-axis, respectively). Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.3.4. Comparison of CD1363 and homologous phage-tail like proteins

Phage-tail like particles are usually composed of an outer part, the contractile sheath, and an inner part, the rigid tube. Based on their structural organization, sheath proteins of phage-tail like particles share several similarities. They are large macromolecules that can be divided into individual domains. One of the best-characterized sheath proteins is that of bacteriophage T4:

protein gp18. It is composed of four domains that are inserted into each other like a Russian doll. Domain IV is inserted between two β-strands of domain III, domain III itself is an insertion into a loop region of domain II and so on. The only crystal structure of gp18 that was published so far is that of a deletion mutant of gp18 (gp18M), which lacks domain I (PDB-ID 3FOA, Aksyuk *et al.*, 2009). The structure and the corresponding topology diagram of gp18M is shown in the top left panel of Fig. 64.

The domain architecture of homologous phage-tail like sheath proteins was compared to that of gp18M (Fig. 64). Several other phage sheath proteins show a similar domain architecture as gp18, but there are some sheath proteins that lack particular domains. They all have in common that their domains are arranged as insertions into the prior domain like a Russian doll. The phage sheath proteins LIN1278 (PDB-ID 3LML, Forouhar *et al.*, 2010, unpublished) and DYS3957 (PDB-ID 3HXL, Forouhar *et al.*, 2010, unpublished) are for example only composed of domains I, II and III and lack domain IV. They seem to depict a truncated form of the bacteriophage sheath protein gp18 and could have evolved from gp18 upon deletion of domain IV. The diffocin sheath protein CD1363 and its homologue, the pyocin sheath protein FIR2 from *Pseudomonas aeruginosa* (PDB-ID 3J9R, Ge *et al.*, 2015), are only composed of domains I and II. As the pyocin sheath protein FIR2 can still assemble into a sheath particle (shown by Cryo-EM studies; Ge e*t al.*, 2015), it can be assumed that domains I and II are the minimal part of bacteriophage-like sheath proteins that are essential for particle formation.

As the diffocin sheath protein CD1363 is only composed of domains I and II, only domains I and II of homologous sheath proteins were used for further analysis. The other domains (III and IV) were neglected for simplification. A structure-based sequence alignment of CD1363, the homologous sheath protein of *Pseudomonas aeruginosa* and the bacteriophage sheath proteins LIN1278 and DSY3957 is shown in Fig. 65 to illustrate sequence similarities. Bold residues indicate conserved amino acids and it can be shown that especially these residues that are organized in secondary structure elements, are conserved among the selected homologous sheath proteins, whereas loop residues are quite variable.

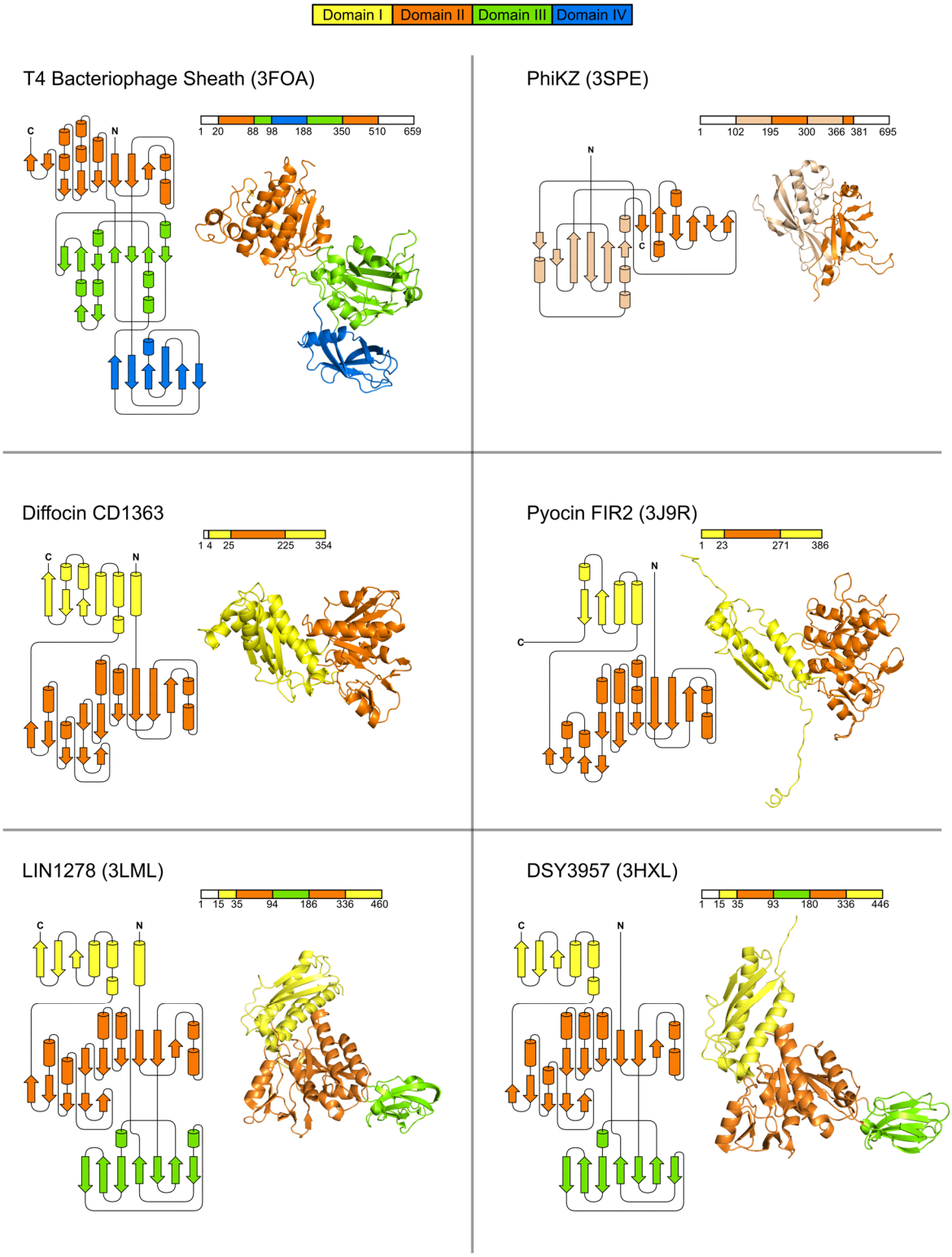

Figure 64: Domain organization of bacteriophage-tail like sheath proteins. Sheath proteins are in general composed of four separate domains, here highlighted by different colors (domain I in yellow, domain II in orange, domain III in green and domain IV in blue). The crystal structure and the corresponding topology diagram of six selected sheath proteins is shown: CD1363 from *Clostridium difficile*, FIR2 from *Pseudomonas aeruginosa* (PDB-ID 3J9R, Ge *et al.*, 2015), T4 Bacteriophage sheath protein gp18 (PDB-ID 3FOA, Aksyuk *et al.*, 2009), sheath protein from bacteriophage PhiKZ (PDB-ID 3SPE, Aksyuk *et al.*, 2011), LIN1278 from *Listeria innocua* (PDB-ID 3LML, unpublished) and DSY3957 from *Desulfitobacterium hafniense* (PDB-ID 3HXL, unpublished). The crystal structures and secondary structure elements are colored according to the domain organization shown at the top of the figure. The domain organization of the full-length amino acid sequence is shown above the respective crystal structure. A white box indicates that this part of the protein is not resolved in the crystal structure or was not included in the crystallized protein construct. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

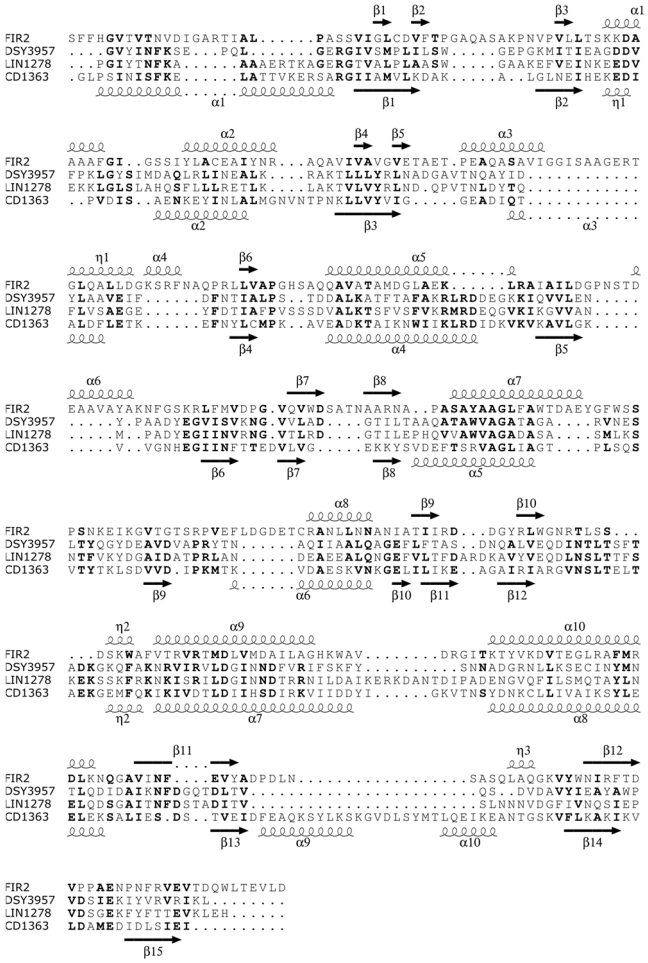

Figure 65: Structure-based sequence alignment of the homologous sheath proteins FIR2 from *Pseudomonas aeruginosa* (PDB-ID 3J9R, Ge *et al.*, 2015), DSY3957 from *Desulfitobacterium hafniense* (PDB-ID 3HXL, unpublished), LIN1278 from *Listeria innocua* (PDB-ID 3LML, unpublished) and CD1363 from *Clostridium difficile*. Only domains I and II were used for this alignment. Domain III of LIN1278 and DSY3957 was left out for clarity. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.3.5. Similarities between the diffocin and pyocin sheath monomer

Among homologous sheath proteins, CD1363 from *Clostridium difficile* shares highest sequence and structural homology to FIR2 from *Pseudomonas aeruginosa*, the component of the contractile sheath of the *Pseudomonas* pyocin particle. Both proteins share a sequence identity of almost 19% and a sequence similarity of 33%. Their structures can be aligned with an r.m.s.d. of 3.9 Å. Although the r.m.s.d. of the structural alignment is quite high, the general topology of CD1363 and FIR2 is highly similar. The three-dimensional structures of both proteins and the corresponding topology diagrams are shown in Fig. 66. Both proteins can be subdivided into two domains, colored in yellow and orange, respectively. Domain II of both sheath proteins contains a central β-sheet composed of six β-strands β1-β6 (one antiparallel and five parallel strands). The long helix α4 of CD1363 that flanks the central β-sheet is divided in two small α-helices α4 and α4* in FIR2. The pyocin sheath protein further possesses an additional helix α5* that is not present in CD1363. The C-terminal part of domain II of CD1363

consists of five short β-strands and two α-helices, similar to that of FIR2 with one β-strand (β9) missing in its structure compared to CD1363. Domain II of CD1363 and FIR2 is hence almost identical with only minor differences in the secondary structure composition.

The main differences between the two homologous sheath proteins can be found in the structural organization of domain I, especially at their N- and C-terminus. The N-terminus of CD1363 is arranged as a long α-helix (α1) that is tightly packed against the rest of the protein. The C-terminus of CD1363 is organized as a β-strand (β14) that interacts with the neighboring β-strand β13. This structural assembly of the CD1363 termini results in a compact packing of the whole protein structure. In contrast to CD1363, both termini of FIR2 are unstructured with two long loops protruding from the rest of the protein. This results in a rather elongated architecture of the FIR2 protein compared to CD1363.

The unstructured protruding terminal regions of FIR2 are essential for the interaction and assembly of the individual FIR2 monomers into the sheath of the pyocin particle. Upon assembly, the termini of FIR2 form short β-strands. The terminal β-strands of the individual monomers interact with each other to form a stable mesh-like interwoven sheath structure of the pyocin particle (Fig. 67). The interaction of the strands is stabilized by π-interactions of neighboring aromatic side chains and hydrogen bonds of neighboring side chains. The assembled pyocin particle is further stabilized by electrostatic attraction of neighboring FIR2 monomers. FIR2 shares a similar electrostatic surface charge distribution as CD1363 (cp. Fig. 60 and Fig. 68). The protruding terminal arms of FIR2 possess a complementary charge, which enables the interaction of the respective arms of FIR2 monomers due to attraction of opposite electrostatic charge. By interaction of several FIR2 monomers, the contractile sheath of the pyocin particle is formed.

The assembly of the pyocin sheath is hence stabilized due to the interaction of the protruding arms of neighboring monomers, the atomic interaction (hydrogen bonds) between adjacent side chains and due to electrostatic attraction of the terminal arms. As these protruding arms are not present in CD1363 monomers, it is unclear how the interaction of the individual monomers upon assembly of the diffocin sheath particle is achieved. One explanation could be that parts of the diffocin sheath protein CD1363 change their conformation upon particle assembly, suggesting that the form crystallized here represents a pre-assembly state with organization of the termini in secondary structure motifs. As the structure of the pyocin sheath monomer FIR2 was solved by cryo-EM studies of the assembled pyocin particle, the structure of the FIR2 monomer depicts the structure and orientation in the assembled form. The obvious structural differences between the termini of both monomers could hence reveal the conformational changes that might be essential for particle assembly. The terminal regions of the diffocin sheath monomer that are tightly packed against the rest of the protein could

therefore undergo conformational changes upon particle assembly and change into protruding arms that interact with each other to form the diffocin sheath, similar to the pyocin particle.

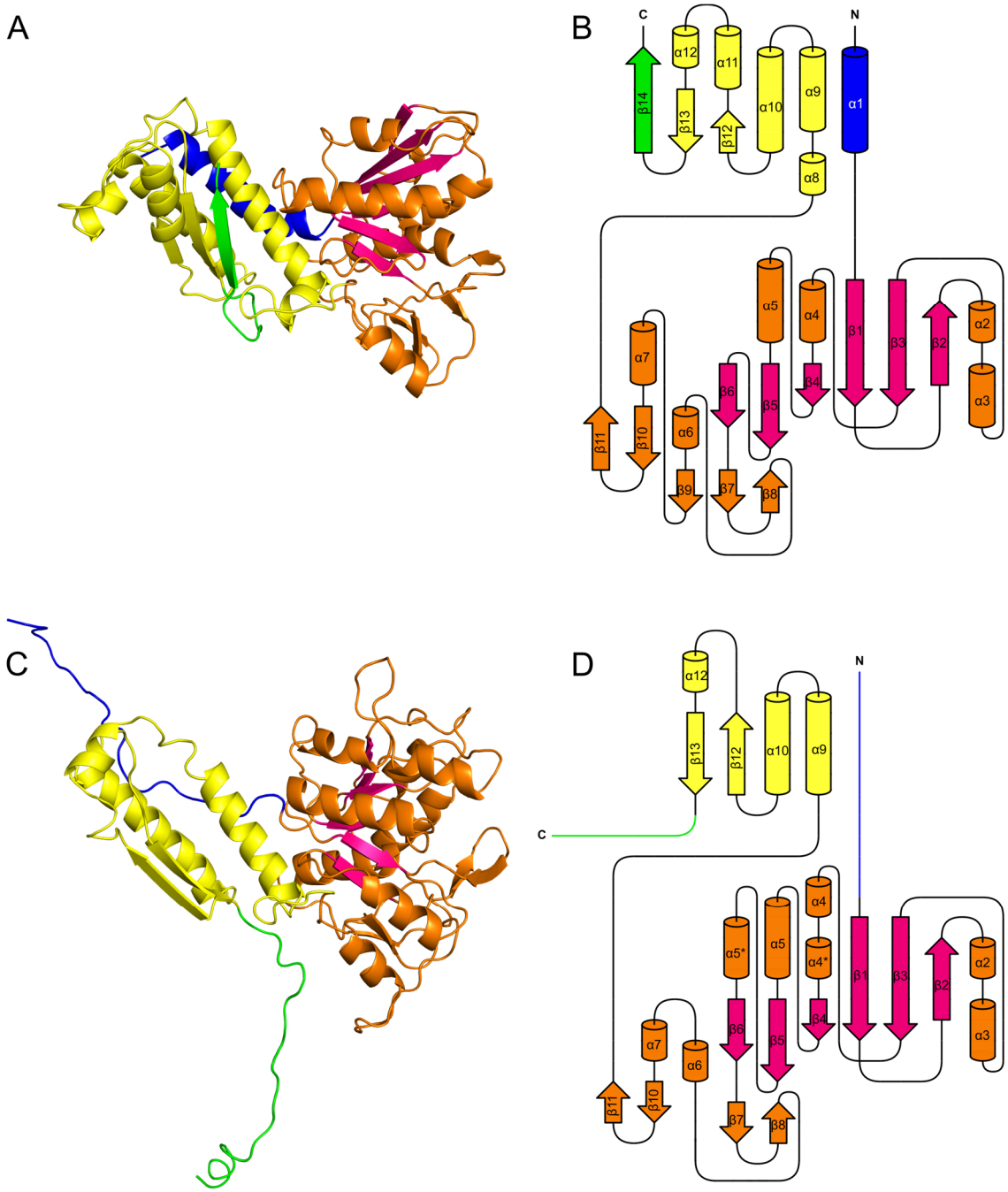

Figure 66: Comparison of the structure and topology of the sheath proteins CD1363 of *Clostridium difficile* and FIR2 of *Pseudomonas aeruginosa* (PDB-ID 3J9R, Ge *et al.*, 2015). The individual domains are colored in yellow and orange, respectively, the central β-sheet of domain II is highlighted in pink, the N-terminus in blue and the C-terminus in green. The topology diagrams are colored according to the crystal structures. α-helices are depicted as cylinders, β-strands as arrows. The secondary structure motifs of CD1363 are numbered consecutively. The elements of FIR2 are numbered according to the homologous parts of CD1363. Helices and strands of FIR2 that are marked with an asterisk (*) confer to secondary structure motifs that are not present in the CD1363 structure. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

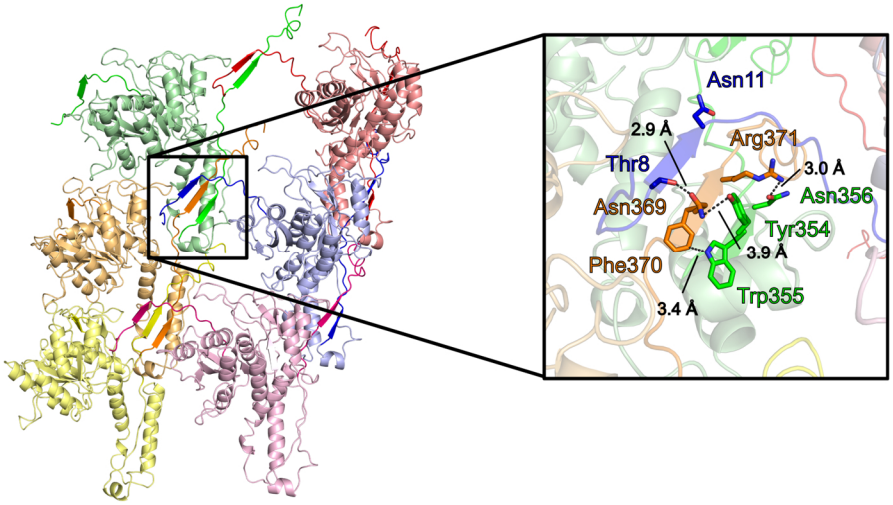

Figure 67: Interaction of FIR2 sheath monomers in the assembled pyocin particle in the elongated form. Six monomers are shown in cartoon representation in light color with the termini in the respective bright color. The right panel illustrates side chains and respective distances between the amino acids that are involved in the interaction of neighboring sheath monomers upon particle assembly. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

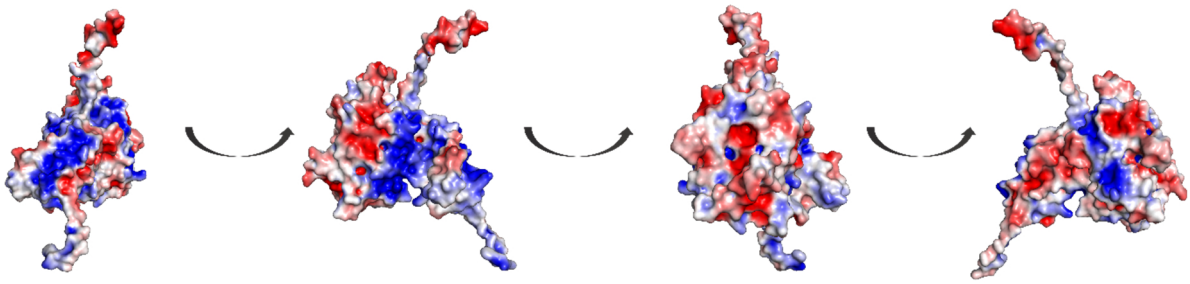

Figure 68: Electrostatic surface of FIR2 from *Pseudomonas aeruginosa*. Individual orientations turned by 90° relative to the prior orientation. The electrostatic surface potential was calculated by *Pymol*. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.3.6. Model of the diffocin sheath particle

Due to the structural similarity of CD1363 and FIR2, it can be assumed that CD1363 monomers interact similar to the pyocin sheath monomers to form a contractile sheath of the diffocin particle. The crystal structure of monomeric CD1363 was aligned on top of the FIR2 monomers of the assembled pyocin sheath (PDB-Codes 3J9Q and 3J9R, Ge *et al.*, 2015) by the Cα-align command in *Pymol* to generate a theoretical model of the assembled diffocin sheath in the contracted as well as in the elongated state. Due to the structural similarity of pyocin and diffocin sheath monomers, the overall architecture of the modelled diffocin particle is therefore almost identical to the assembled pyocin sheath. The individual CD1363 monomers do not clash with each other in the elongated, as well as in the contracted diffocin model. The final models of the diffocin particle in the pre- and post-contraction state are shown in Fig. 69. Because CD1363 was most probably crystallized in its “pre-assembly form”, no interaction of the individual CD1363 monomers can be observed in the model of the diffocin particle.

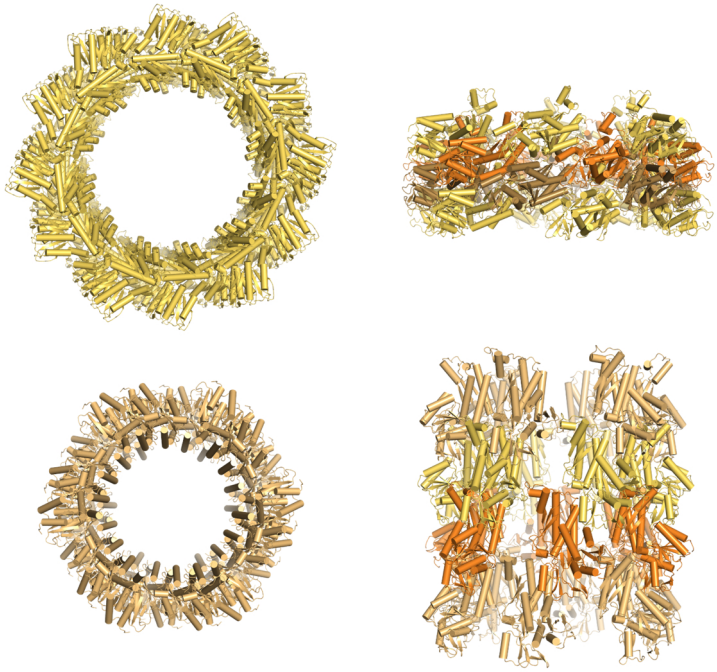

Figure 69: A model of the diffocin particle in its pre- (bottom) and post-contraction state (top) based on the pyocin structure in pre- (PDB-Code 3J9R) and post-contraction state (PDB-Code 3J9Q). The diffocin sheath in the post-contraction state (top panel) is colored in light orange and shown from the top (left) and from the side (right). The diffocin sheath in the pre-contraction state (lower panel) is colored in light yellow and shown from the top (left) and the side perspective (right). Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

The assembled pyocin particle as well as the model of the diffocin particle are arranged in a six-start helix, which is composed of six sheath monomers. The sheath particle is composed of several discs of each six CD1363 monomers that are packed on top of each other. It is assumed that the individual CD1363 monomers in the pre- and post-contraction state of the diffocin particle share the same conformation as compared to the pyocin particle (Ge *et al.*, 2015). Upon contraction, the pyocin sheath monomers are tilted relative to their pre-contraction orientation. Therefore, the distance between the individual discs is decreased and the diameter of the sheath particle is increased. The monomers in the assembled pyocin particle are oriented in a way that domain I points to the inside of the sheath particle and domain II is directed to the outside of the particle. Due to the structural similarity between CD1363 and homologous sheath proteins, it is assumed that CD1363 monomers are arranged in the same orientation as e.g. in the pyocin particle and that the mechanism of contraction of the diffocin sheath is similar to that of the pyocin sheath.

Similar to the pyocin particle, the model of the diffocin sheath particle shows a predominantly negatively charged surface on both the outside and on the inside (Ge *et al.*, 2015; Fig. 70B).

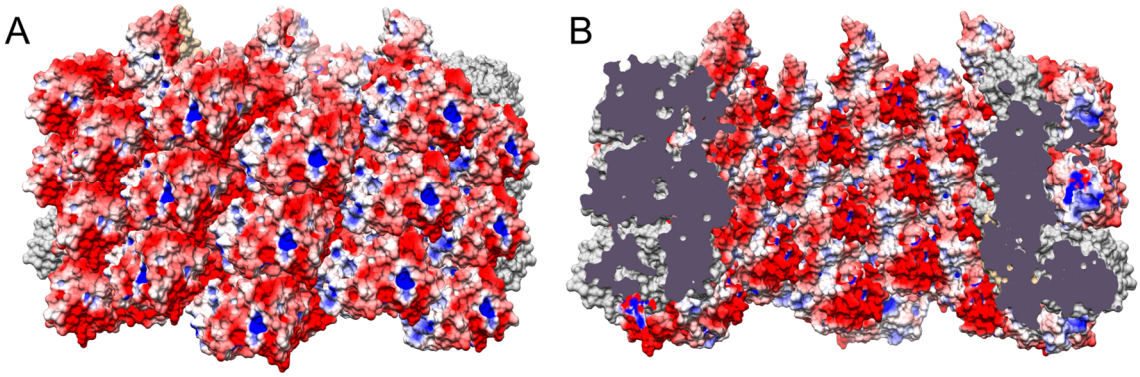

Figure 70: The model of the contracted diffocin sheath is colored according to its electrostatic potential and is depicted from the side view (A) and from the inside (B) in a clipped representation. Red patches illustrate negative charge, blue patches indicate positive charge. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.4. The diffocin tube protein CD1364

The diffocin tube protein CD1364 is a 16 kDa protein and is annotated as a putative phage XkdM-like protein. Due to homology to phage tail-like tube proteins and its genetic context, it is assumed that it is the tube component of the diffocin particle.

8.4.1. From gene to structure

The gene encoding CD630_13640 (short CD1364) was amplified from genomic *C. difficile* CD630 DNA and was cloned in a pOPIN-M vector. Expression and purification was performed analogously to CD1363 (chapter 8.3.1). The final size exclusion chromatography chromatogram and the corresponding SDS-PAGE of pure CD1364 are shown in Fig. 71. Pure protein was used for crystallization experiments using commercially available screens. Crystals were obtained in some conditions after a few days and exemplary crystals of native CD1364 are depicted in Fig. 72. Native datasets were collected at BESSY (Berlin, Germany; Müller *et al.*, 2015) and the structure was solved by Molecular Replacement using a truncated model of chain A of P54332 from *Bacillus subtilis* (PDB-Code 2GUJ) to 1.5 Å resolution. The protein was crystallized by Dr. Emerich-Mihai Gazdag, structure solution was performed with the help of Dr. Jan Pippel. CD1364 crystallized in space group $P4_32_12$ with one molecule in the asymmetric unit (solvent content 44.86%, Matthews coefficient 2.23). Statistics of data collection and refinement are shown in table 29 and table 30, respectively. The overall structure of monomeric CD1364 in cartoon representation and the corresponding topology diagram are illustrated in Fig. 73.

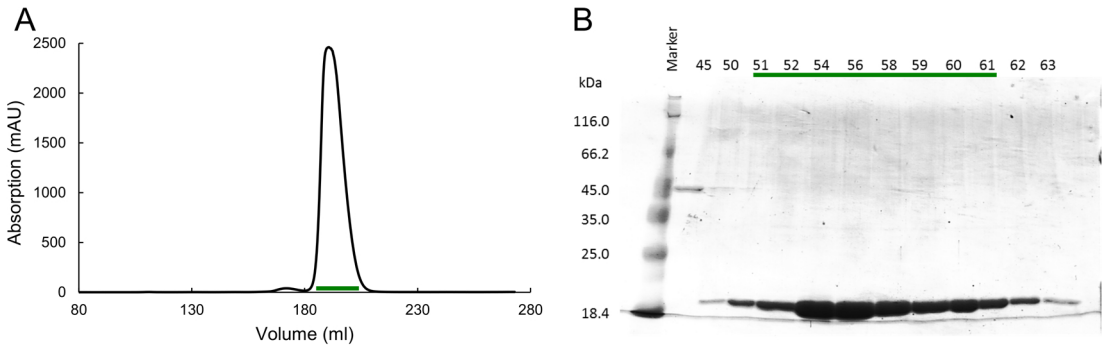

Figure 71: Size exclusion chromatography of CD1364 (A) and corresponding SDS-PAGE (B) of eluted fractions. A protein standard is shown in the first lane of the SDS-PAGE for molecular weight comparison of CD1364 (theoretical molecular weight 16 kDa). The green bar indicates the protein fractions that were used for crystallization setups. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

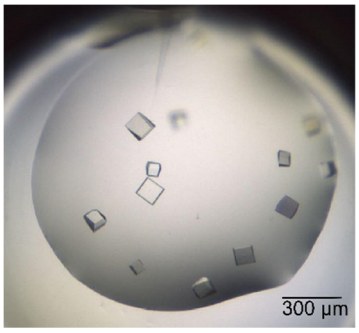

Figure 72: Exemplary protein crystals of native CD1364 (Crystallization condition: 20% (w/v) PEG3350, 0.2 M ammoniumchloride, 20 mg/ml CD1364). Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

Table 29: Data collection statistics for native CD1364 crystal.

Dataset	**CD1364**
Wavelength (Å) / beamline	0.918 Å / BESSY 14.1
Resolution range (Å)	34.42-1.50 (1.55-1.50)
Space group	$P4_32_12$
Unit cell parameters (Å)	36.3 36.3 216.7
(°)	90 90 90
Mosaicity (°)	0.10
Total No. of measured reflections	611841 (30831)
Unique reflections	24674 (1193)
Multiplicity	24.8 (25.8)
Mean I/σ(I)	19.6 (1.5)
Completeness (%)	100 (100)
R_{meas} (%)	8.5 (257.2)
$R_{p.i.m.}$ (%)	2.3 (68.7)
CC1/2	100 (62.8)

Table 30: Refinement statistics for native CD1364.

Dataset	**CD1364**

Resolution Range (Å)	35.83-1.50 (1.57-1.50)
R_{work} (%)	19.86
R_{free} (%)	21.91
No. of non-H atoms	
Protein	1079
Water	128
R.m.s. deviation	
Bonds (Å)	0.006
Angles (°)	0.698
Average B factors (Å^2)	
Protein	36
Water	41
All atoms	37
Ramachandran plot	
Favored regions (%)	98.5
Outliers (%)	1.5
MolProbity score	1.28

8.4.2. The crystal structure of CD1364

Monomeric CD1364 consists of a central seven-stranded β-barrel with alternating parallel and antiparallel β-strands. The barrel is flanked by two α-helices on each side, which cap the top and the bottom part of the central β-barrel.

Analysis of the CD1364 crystal structure regarding conservation of amino acids reveals that especially the residues of the central β-barrel are highly conserved among homologous proteins. The flanking α-helices are more variable in their amino acid composition (Fig. 74). The central β-barrel is conserved among bacteriophage-like tube proteins (chapter 8.4.4 for further analysis), while the helices surrounding the β-barrel of CD1364 are unique in the diffocin tube protein and cannot be found in homologous tube proteins.

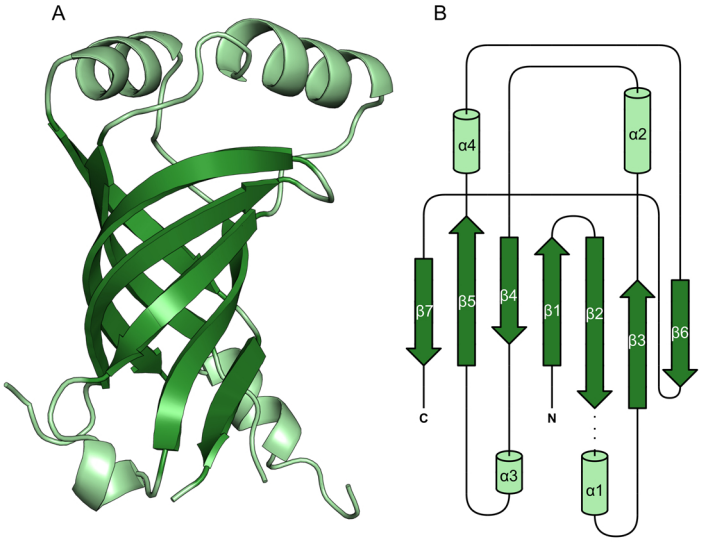

Figure 73: Crystal structure of monomeric CD1364 in cartoon representation (A) and as topology diagram (B). The central β-barrel is colored in dark green, surrounding α-helices are shown in light green. In (B), α-helices are illustrated as cylinders and β-strands as arrows. The secondary structure elements are numbered consecutively. The dashed line indicates a region that could not be built due to unresolved electron density. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

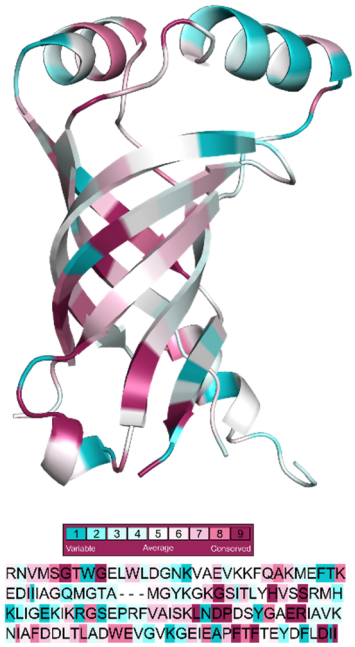

Figure 74: Conservation of CD1364. The crystal structure of monomeric CD1364 (top) and its amino acid sequence (bottom) is depicted colored according to conservation score with 1 (cyan) highlighting variable and 9 (dark red) illustrating conserved residues. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.4.3. SAXS analysis of CD1364

The oligomeric state of CD1364 in solution was analyzed by SAXS experiments. Best data were measured at BM29, ESRF (Grenoble, France; Pernot *et al.*, 2013) for 1.25 mg/ml CD1364 in 20 mM HEPES pH 7.5, 150 mM NaCl. The fit of the experimental SAXS data and theoretical data for monomeric CD1364 is shown in Fig. 75. The fit with theoretical data for

monomeric CD1364 is characterized by a lower *X*-value (0.71) compared to the fit for dimeric CD1364 (*X*-value of 21.33; fit not shown). Due to the lower *X*-value and hence the better fit of experimental and theoretical data, it could be shown by SAXS experiments that CD1364 is monomeric in solution. The SAXS envelope was calculated for monomeric CD1364 and the crystal structure was fit into the envelope (Fig. 76).

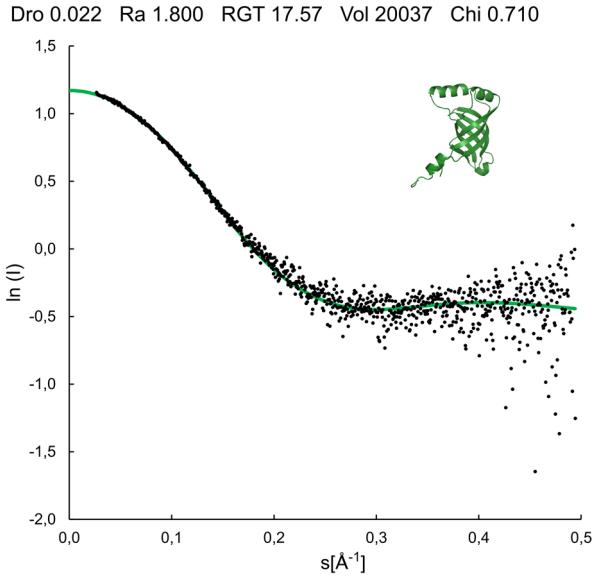

Figure 75: Fit of experimental SAXS data (black dots) with calculated curves for monomeric (A) and dimeric CD1364 (B) (green curves). Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

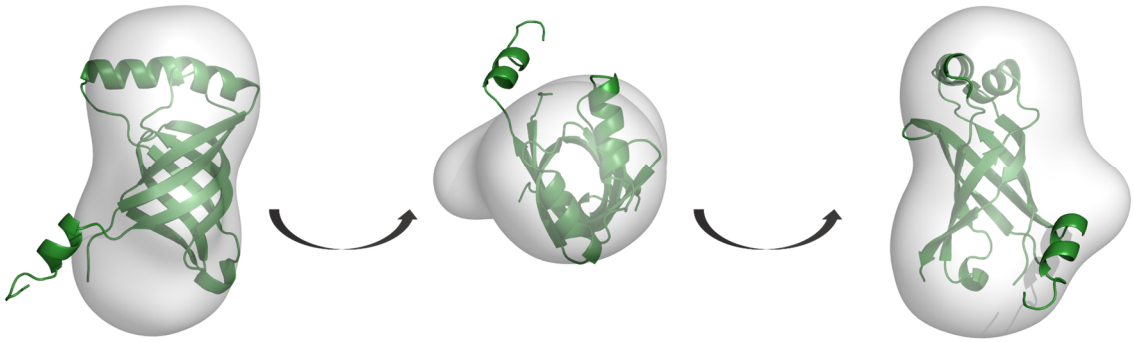

Figure 76: Crystal structure of monomeric CD1364 (green cartoon representation) fitted into the calculated SAXS envelope for monomeric CD1364. Individual figures are turned by 90° around x- and y-axis respectively. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

The fit of the crystal structure into the SAXS envelope reveals that most of the protein in the crystal state is arranged in the same orientation as in solution. One exception is the region involving the small helix α1 of the crystal structure of CD1364, which could not be fit into the SAXS envelope. This, together with the lack of electron density in this region, indicates that this part of the protein is highly flexible and can adopt different orientations in the crystal, as well as in solution. The high flexibility of this region is further indicated by high B-factors in the crystal structure.

8.4.4. Comparison of CD1364 with homologous tube proteins

Homologous proteins of the diffocin tube protein CD1364 are e.g. the pyocin tube protein FIIR2 from *Pseudomonas aeruginosa*, the T4 bacteriophage tube protein gp19 or the Hcp tube protein of the T6SS system of *Vibrio cholera*. A comparison of the topology and three-dimensional structure of these proteins is illustrated in Fig. 77. Secondary structure motifs that are homologous to CD1364 are colored in the same color as in the CD1364 structure, which is depicted in the top panel of the figure.

All of these selected tube proteins have in common that they are arranged as a central β-barrel, which is composed of antiparallel β-strands. The number of strands involved in barrel formation differs between the individual tube proteins, together with the secondary structure elements that surround the β-barrel.

Based on the architecture of the monomer, CD1364 shows highest similarity to the monomer of XkdM of *Bacillus subtilis*/bacteriophage phi812K1-420 (PDB-ID 5LI2, Novacek *et al*., 2016), as both monomers can be aligned with an r.m.s.d. of 1.895 Å. The only difference in their secondary structure composition are two additional short α-helices (α1 and α3) in the CD1364 structure that are not present in XkdM. Some parts of the XkdM structure are not resolved due to low resolution and high flexibility, which is indicated by a dashed line in the topology diagram in Fig. 77. The other homologous tube proteins share a similar topology as CD1364, but have several missing or additional secondary structure elements compared to CD1364. The alignment of these tube proteins with CD1364 was hence not as good as the alignment with the XkdM tube protein.

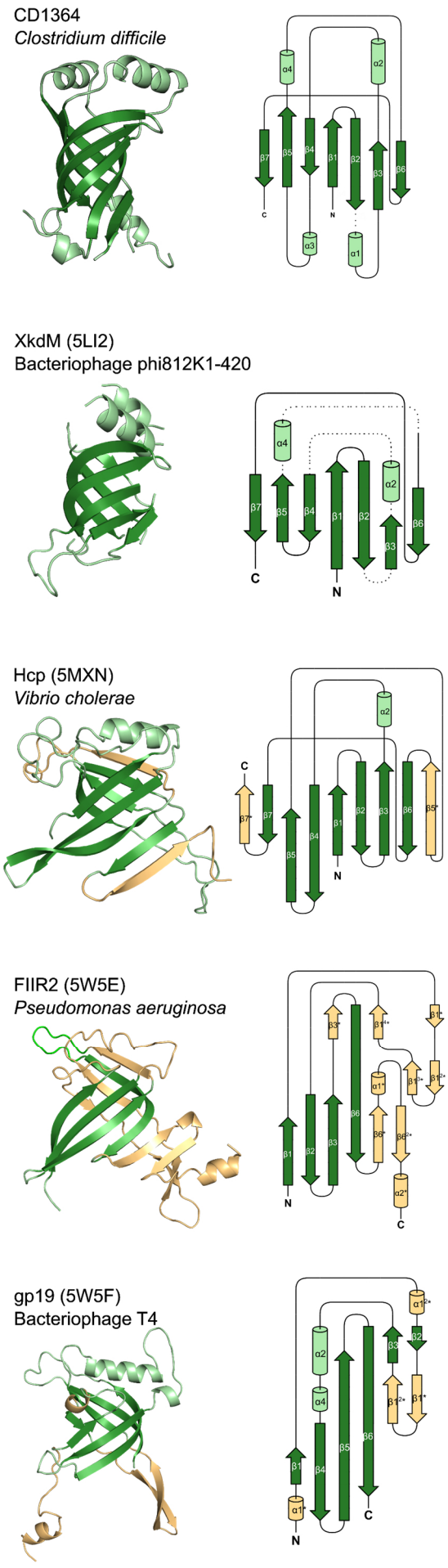

Figure 77: Structural comparison of the diffocin tube protein CD1364 from *Clostridium difficile*, tube protein XkdM (PDB-ID 5LI2, Novacek *et al.*, 2016) from bacteriophage phi812K-420, T6SS tube protein Hcp (PDB-ID 5MXN, Wang *et al.*, 2017) from *Vibrio cholerae*, pyocin tube protein FIIR2 (PDB-ID 5W5E, Zheng *et al.*, 2017) from *Pseudomonas aeruginosa* and tube protein gp19 (PDB-ID 5W5F, Zheng *et al.*, 2017) from bacteriophage T4. The crystal structure of the tube proteins is shown in the left panel, the corresponding topology diagram in the right panel. Each structure was colored according to the color code of CD1364. Here, the central β-barrel is colored in dark green and surrounding α-helices in pale green. Secondary structure elements that could be aligned to CD1364 were colored accordingly to CD1364. Additional elements that are not present in the CD1364 structure were colored in light orange. In the topology diagrams, α-helices are depicted as cylinders and β-sheets as arrows. The secondary structure elements of CD1364 were numbered consecutively. The elements of the other tube proteins were numbered according to the respective aligned secondary structure element of CD1364. Additional elements that are not present in CD1364 were marked with a star. Regions that are not resolved in the crystal structure are indicated by a dashed line in the topology diagram. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.4.5. Model of the diffocin tube particle

All tube proteins shown in Fig. 77 have in common that they assemble in several hundred copies into a hollow tube particle. These tube particles are arranged as a six-start helix with six tube monomers arranged in a circular manner. Several of such discs are packed on top of each other to build the tube particle similar to the sheath particle. The diffocin tube protein CD1364 was aligned on top of the homologous tube particles to compare their structural arrangement and to derive an idea about the orientation of the CD1364 monomer in the assembled diffocin particle.

Cryo-EM studies revealed the orientation and architecture of the bacteriophage phi812K1-420 (PDB-ID 5LI2, Novacek *et al*., 2016) sheath-tube complex at low resolution (6.8 Å). The crystal structure of the CD1364 monomer was aligned on top of the tube monomers, but the aligned CD1364 monomers clash with each other due to the presence of the two additional α-helices of CD1364. In detail, the CD1364 helix α4 clashes with strand β1 of the neighboring CD1364 monomer. CD1364 monomers were also aligned on top of the other assembled tube particles, but in each model, some secondary structure elements of CD1364 monomers clashed with each other (α3 and α4 for the fit with Hcp from *Vibrio cholerae*, α2 and α3 for the model based on FIIR2 from *Pseudomonas aeruginosa*).

A model of the diffocin tube particle without serious clashes could be generated by alignment of the CD1364 monomers on top of the assembled tube proteins gp19 of bacteriophage T4 (PDB-ID 5W5F, Zheng *et al*., 2017). The central β-barrel is arranged in a way that the axis of the barrel lies perpendicular to the assembled tube plane. The additional β-strands of the T4 tube protein gp19, which are not present in the CD1364 structure, are located on top of the barrel and point to the inside of the tube. They could be involved in stabilization of the interaction between the neighboring monomers. The additional short α-helix of gp19, which cannot be found in the CD1364 structure, is located on the other side opposite of the two additional β-strands and could hence facilitate the interaction with the neighboring disc that is packed on top of the prior one. As these secondary structure elements are not present in CD1364, the interaction of CD1364 monomers and the stabilization of the assembled diffocin tube remains unclear. All tube particles of the other homologous bacteriophage-like proteins share the same orientation of the central β-barrel, which lies perpendicular to the tube axis with flanking secondary structure elements that stabilize the interaction between neighboring tube monomers. The diffocin tube model (Fig. 78) based on alignment on top of the bacteriophage T4 tube shows that the CD1364 helices α1 and α2 (colored in orange and yellow in the diffocin model) point away from the rest of the protein and could hence be involved in interaction of the neighboring discs. Thus, these helices could stabilize the assembly of the tube particle. One disc of the diffocin tube model is shown in the left and middle panel of Fig. 78 from the top and side view, respectively. Helices that could be involved in interaction of the

individual discs are highlighted in yellow and orange. The model of a diffocin tube composed of three discs is shown in the right panel of Fig. 78.

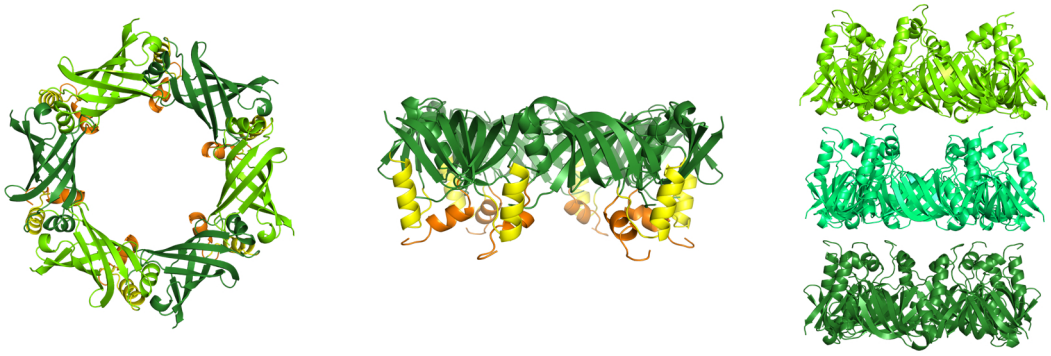

Figure 78: Diffocin tube model (one disc) based on homology to the assembled tube particle of bacteriophage T4 (PDB-ID 5W5F, Zheng *et al*., 2017). Diffocin CD1364 monomers are shown in green with helices that are possibly involved in interaction with neighboring discs in yellow and orange, respectively. The diffocin tube disc is shown from the top (left panel) and from the side view (middle panel). A model of an assembled diffocin tube composed of three discs is shown in the right panel. Figure is adapted from Schwemmlein *et al*. (Front. Microbiol., 2018).

Similar to the sheath protein CD1363, the CD1364 monomer shows a complementary distribution of positively and negatively charged surface patches. A positively charged patch on one side of the monomer lies opposite to a negatively charged patch on the other side of the protein (Fig. 79A). This enables the interaction of CD1364 monomers due to electrostatic attraction. In the diffocin tube model, these oppositely charged regions interact with each other, creating a stable interaction of the tube monomers and could also be involved in self-assembly of the monomers into the diffocin particle. The outside of the tube model shows an even distribution of positive and negative charge. These charged patches could interact with oppositely charged patches of the sheath particle to facilitate a stable interaction between the sheath and tube particle. The inside of the diffocin tube model also shows an even distribution of positively and negatively charged patches. It is assumed that the diffocin particle is a channel used for ion transport similar to the pyocin particle (Ge *et al*., 2015). Due to the even charge distribution, ions could easily be transported across this channel.

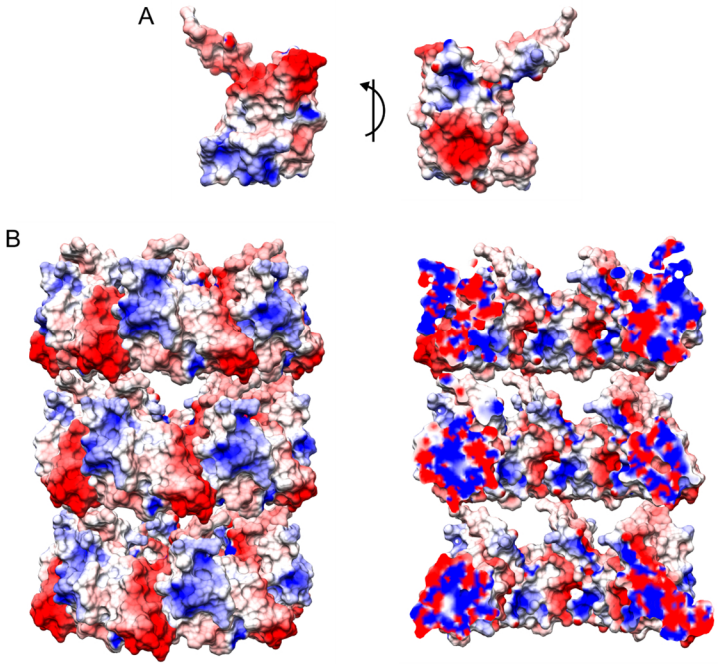

Figure 79: Electrostatic surface charge distribution of CD1364. The surface charge of the CD1364 monomer is shown in two different orientations (A) projected by 180° rotation around the illustrated axis. The electrostatic surface charge of the modelled diffocin tube particle is illustrated in (B) from the outside (left) and from the inside (right). Red color indicates negative charge, blue color illustrates positive charge. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

8.4.6. Model of the assembled diffocin particle

The homologous pyocin particle was analyzed by Cryo-EM studies in its elongated as well as in the contracted state. In the elongated state, the electron density for sheath and tube was resolved to 3.5 Å and the structure of the individual sheath and tube monomers could be modelled into the density map. In contrast, the density for the pyocin tube was not resolved in the contracted state of the pyocin particle. Only the structure of the sheath monomers could be modelled into the density map of the contracted particle to 3.9 Å resolution (Ge *et al.*, 2015). A model of the fully assembled diffocin particle composed of sheath and tube was therefore generated based on the pyocin particle in its elongated form. This model is shown in Fig. 80. Some minor clashes are present between the diffocin sheath and tube particle, i.e. the clash of the loop between β1 and β2 of CD1364 (tube) with helix α10 of CD1363 (sheath).

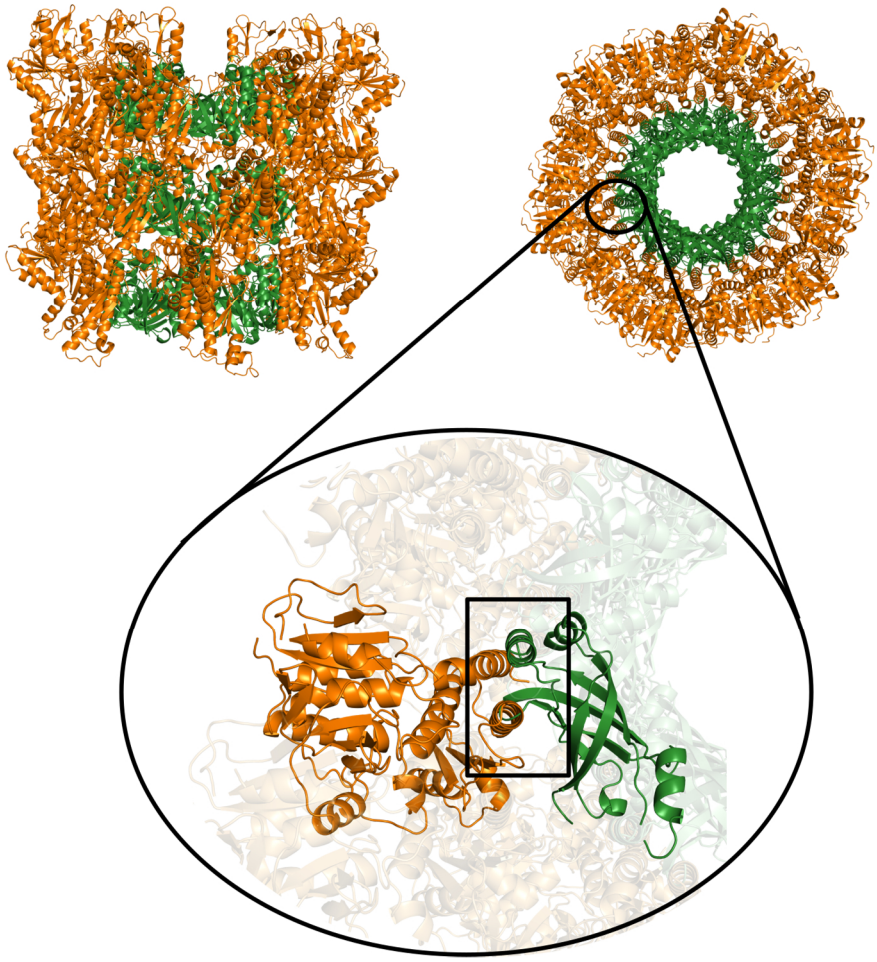

Figure 80: Model of the diffocin particle in the elongated state based on the architecture of the pyocin particle (PDB-ID 3J9Q, Ge *et al.*, 2015). The diffocin particle is composed of an outer sheath (orange) and an inner tube (green). It is shown from the side (left) and from the top (middle panel). The right figure illustrates the clashes between sheath and tube monomer in the diffocin particle model. Figure is adapted from Schwemmlein *et al.* (Front. Microbiol., 2018).

In summary, the models of the assembled diffocin particle shown above are only theoretical models based on homology of the crystal structures of the monomeric diffocin sheath and tube protein to homologous bacteriophage tail-like proteins. Due to some clashes in the prepared models of the diffocin particle, it is probable that the diffocin monomers have to undergo slight conformational changes upon particle assembly. The exact structure and architecture of the assembled diffocin particle needs additional experiments, such as cryo-EM studies similar to the pyocin particle to reveal the orientation and assembly of the diffocin particle in nature.

9. Conclusions and Outlook

This thesis was subdivided into three different projects: the toxins of *Clostridium difficile*, essential proteins involved in growth or sporulation of the pathogen and the diffocin proteins. Aim of all projects was the structural analysis of the respective virulence factors and essential proteins to gain a deeper understanding of their structure and function. The detailed structural characterization of the so far uncharacterized proteins should reveal ideas about their potential molecular function and facilitate the comparison with structurally homologous proteins.

The main virulence factor TcdB from *Clostridium difficile* was already characterized to some extent by other groups regarding structure and molecular function of its individual domains. The only domain that was not characterized so far was the intermediate domain, which is involved in translocation and pore formation. SAXS analysis of this domain was performed in this thesis and revealed that the TcdB translocation domain shares the same overall shape and oligomeric state as the TcdA translocation domain. Nevertheless, determination of a crystal structure of the TcdB translocation domain is necessary to analyze its molecular structure and to draw a detailed comparison to the TcdA translocation domain at atomic level. The exact mechanism of pore formation and translocation of both homologous toxins across the endosomal membrane are also not completely understood. Hence, further analysis of toxin translocation has to be performed to understand this part of the toxin mechanism, especially in terms of designing potential drugs targeting this process of toxin translocation.

The second project involved the characterization of two essential proteins involved in growth or sporulation of *Clostridium difficile*. Although the structural analysis of CD1219 and CD1823 by X-ray crystallography and SAXS revealed some ideas about their molecular function, these hypotheses have to be validated by further experiments. A comparison with structurally homologous proteins revealed that CD1219 (determined by X-ray crystallography to 1.8 Å resolution) shows strong similarity to GTPases and that CD1823 (determined by X-ray crystallography to 1.9 Å resolution) could possibly be involved in H_2O_2 stress response. Future investigations will include a detailed biochemical analysis of CD1219 regarding ligand binding and potential GTPase activity. Several potential ligands have to be tested towards their binding specificity for CD1219. A broad TDSA of CD1219 could be performed to test the stabilizing or destabilizing effect of several potential ligands on the protein. Ligands that show a stabilizing effect can then be subjected to binding studies by MST or ITC measurements. Furthermore, soaking or co-crystallization experiments can be performed with potential ligands to determine ligand-bound structures of CD1219. Such ligand structures aid to understand the binding mechanism of CD1219. Additional biochemical analysis should be performed to reveal the molecular function of CD1823 and its potential role in H_2O_2 stress response. Transcriptomic or metabolomics analysis can thereby help to identify the molecular function of CD1823.

A third project dealt with the characterization of diffocin proteins, bacteriophage tail-like bacteriocins of *Clostridium difficile*. The determination of the first crystal structure of the diffocin sheath protein CD1363 (1.9 Å resolution) and of the tube protein CD1364 (1.5 Å resolution) in this thesis revealed important details about the molecular components of the diffocin tail. Comparison with homologous phage-tail like particles led to a generation of a theoretical model of the diffocin particle. Nevertheless, an analysis of the assembled native diffocin particle is essential to understand its molecular organization and architecture. Cryo-EM analysis of the diffocin particle would be an important technique to yield a structure of the assembled particle. To summarize, it could be shown in this thesis that structural analysis of proteins using X-ray crystallography can reveal ideas about the potential function of these proteins, but structural analysis alone is not sufficient to characterize the molecular function of a protein. Further additional biochemical and biophysical experiments always have to be performed to prove the exact function of the respective proteins. Nevertheless, the novel protein structures that have been determined in this thesis revealed important structural details of the analyzed proteins and serve as a basis for the design of future experiments for a functional analysis of these proteins.

10. References

Ackermann, H.-W. (2003). Bacteriophage observations and evolution. Res. Microbiol. *154*, 245–251.

Adams, C.M., Eckenroth, B.E., Putnam, E.E., Doublié, S., and Shen, A. (2013). Structural and functional analysis of the CspB protease required for *Clostridium* spore germination. PLoS Pathog. *9*, e1003165.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr *66*, 213–221.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.-W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W., et al. (2011). The Phenix software for automated determination of macromolecular structures. Methods *55*, 94–106.

Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., and Adams, P.D. (2010). Joint X-ray and neutron refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *66*, 1153–1163.

Aksyuk, A.A., Leiman, P.G., Kurochkina, L.P., Shneider, M.M., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2009). The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. EMBO J. *28*, 821–829.

Albesa-Jové, D., Bertrand, T., Carpenter, E.P., Swain, G.V., Lim, J., Zhang, J., Haire, L.F., Vasisht, N., Braun, V., Lange, A., et al. (2010). Four distinct structural domains in *Clostridium difficile* toxin B visualized using SAXS. J. Mol. Biol. *396*, 1260–1270.

Ali, S., Moore, G., and Wilson, A.P.R. (2011). Spread and persistence of *Clostridium difficile* spores during and after cleaning with sporicidal disinfectants. J. Hosp. Infect. *79*, 97–98.

Alteri, C.J., and Mobley, H.L.T. (2016). The Versatile Type VI Secretion System. Microbiol Spectr *4*.

Anand, B., Verma, S.K., and Prakash, B. (2006). Structural stabilization of GTP-binding domains in circularly permuted GTPases: implications for RNA binding. Nucleic Acids Res. *34*, 2196–2205.

Arcus, V. (2002). OB-fold domains: a snapshot of the evolution of sequence, structure and function. Curr. Opin. Struct. Biol. *12*, 794–801.

Arisaka, F., Yap, M.L., Kanamaru, S., and Rossmann, M.G. (2016). Molecular assembly and structure of the bacteriophage T4 tail. Biophys Rev *8*, 385–396.

Asherie, N. (2004). Protein crystallization and phase diagrams. Methods *34*, 266–272.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344-350.

Awad, M.M., Johanesen, P.A., Carter, G.P., Rose, E., and Lyras, D. (2014). *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. Gut Microbes *5*, 579–593.

Baban, S.T., Kuehne, S.A., Barketi-Klai, A., Cartman, S.T., Kelly, M.L., Hardie, K.R., Kansau, I., Collignon, A., and Minton, N.P. (2013). The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. PLoS ONE *8*, e73026.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. PNAS *98*, 10037–10041.

Barra-Carrasco, J., and Paredes-Sabja, D. (2014). *Clostridium difficile* spores: a major threat to the hospital environment. Future Microbiol *9*, 475–486.

Bartlett, J.G., Moon, N., Chang, T.W., Taylor, N., and Onderdonk, A.B. (1978). Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. Gastroenterology *75*, 778–782.

Beilhartz, G.L., Tam, J., and Melnyk, R.A. (2015). Small Molecules Take A Big Step Against *Clostridium difficile*. Trends Microbiol. *23*, 746–748.

Bender, K.O., Garland, M., Ferreyra, J.A., Hryckowian, A.J., Child, M.A., Puri, A.W., Solow-Cordero, D.E., Higginbottom, S.K., Segal, E., Banaei, N., et al. (2015). A small-molecule antivirulence agent for treating *Clostridium difficile* infection. Sci Transl Med *7*, 306ra148.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000a). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

Berman, H.M., Bhat, T.N., Bourne, P.E., Feng, Z., Gilliland, G., Weissig, H., and Westbrook, J. (2000b). The Protein Data Bank and the challenge of structural genomics. Nat. Struct. Biol. *7 Suppl*, 957–959.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., et al. (2002). The Protein Data Bank. Acta Crystallogr. D Biol. Crystallogr. *58*, 899–907.

Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc. *129*, 5656–5664.

Bhattacharjee, D., Francis, M.B., Ding, X., McAllister, K.N., Shrestha, R., and Sorg, J.A. (2015). Reexamining the Germination Phenotypes of Several *Clostridium difficile* Strains Suggests Another Role for the CspC Germinant Receptor. J. Bacteriol. *198*, 777–786.

Bingel-Erlenmeyer, R., Olieric, V., Grimshaw, J. P. A., Gabadinho, J., Wang, X., Ebner, S. G., Isenegger, A., Schneider, R., Schneider, J., Glettig, W., Pradervand, C., Panepucci, E. H., Tomizaki, T., Wang, M., and Schulze-Briese, C. (2011). SLS Crystallization Platform at Beamline X06DA—A Fully Automated Pipeline Enabling in Situ X-ray Diffraction Screening. Cryst. Growth Des., *11*, 916–923.

Bönemann, G., Pietrosiuk, A., and Mogk, A. (2010). Tubules and donuts: a type VI secretion story. Mol. Microbiol. *76*, 815–821.

Bourne, H.R., Sanders, D.A., and McCormick, F. (1990). The GTPase superfamily: a conserved switch for diverse cell functions. Nature *348*, 125–132.

Bourne, H.R., Sanders, D.A., and McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. Nature *349*, 117–127.

Bragg, W.H., A, M., S, F.R., Bragg, W.L., and A, B. (1913). The reflection of X-rays by crystals. Proc. R. Soc. Lond. A *88*, 428–438.

Brennan, C.A., and Platt, T. (1991). Mutations in an RNP1 consensus sequence of Rho protein reduce RNA binding affinity but facilitate helicase turnover. J. Biol. Chem. *266*, 17296–17305.

Brito, G.A.C., Sullivan, G.W., Ciesla, W.P., Carper, H.T., Mandell, G.L., and Guerrant, R.L. (2002a). *Clostridium difficile* toxin A alters in vitro-adherent neutrophil morphology and function. J. Infect. Dis. *185*, 1297–1306.

Brito, G.A.C., Fujji, J., Carneiro-Filho, B.A., Lima, A.A.M., Obrig, T., and Guerrant, R.L. (2002b). Mechanism of *Clostridium difficile* toxin A-induced apoptosis in T84 cells. J. Infect. Dis. *186*, 1438–1447.

Browning, C., Shneider, M.M., Bowman, V.D., Schwarzer, D., and Leiman, P.G. (2012). Phage pierces the host cell membrane with the iron-loaded spike. Structure *20*, 326–339.

Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. *41*, W349-357.

Bulheller, B.M., Rodger, A., and Hirst, J.D. (2007). Circular and linear dichroism of proteins. Phys Chem Chem Phys *9*, 2020–2035.

Burger, S., Tatge, H., Hofmann, F., Genth, H., Just, I., and Gerhard, R. (2003). Expression of recombinant *Clostridium difficile* toxin A using the Bacillus megaterium system. Biochem. Biophys. Res. Commun. *307*, 584–588.

Burkhardt, A., Pakendorf, T., Reime, B., Meyer, J., Fischer, P., Stübe, N., Panneerselvam, S., Lorbeer, O., Stachnik, K., Warmer, M., et al. (2016). Status of the crystallography beamlines at PETRA III. Eur. Phys. J. Plus *131*, 56.

Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D., and Aktories, K. (1998). A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. Journal of Biological Chemistry *273*, 19566–19572.

Cadena, J., Thompson, G.R., Patterson, J.E., Nakashima, B., Owens, A., Echevarria, K., and Mortensen, E.M. (2010). Clinical predictors and risk factors for relapsing *Clostridium difficile* infection. Am. J. Med. Sci. *339*, 350–355.

Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Hartsch, T., Wimberly, B.T., and Ramakrishnan, V. (2001). Crystal structure of an initiation factor bound to the 30S ribosomal subunit. Science *291*, 498–501.

Carvajal, A.I., Vallejos, G., Komives, E.A., Castro-Fernández, V., Leonardo, D.A., Garratt, R.C., Ramírez-Sarmiento, C.A., and Babul, J. (2017). Unusual dimerization of a BcCsp mutant leads to reduced conformational dynamics. FEBS J. *284*, 1882–1896.

Centers for Disease Control and Prevention (CDC). Antibiotic resistance threats in the United States, 2013. (2013). Available from: http://www.cdc.gov/drugresistance/ threat-report-2013/pdf/ar-threats-2013-508.pdf

Cerquetti, M., Molinari, A., Sebastianelli, A., Diociaiuti, M., Petruzzelli, R., Capo, C., and Mastrantonio, P. (2000). Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. Microb. Pathog. *28*, 363–372.

Chaikam, V., and Karlson, D. (2008). Functional characterization of two cold shock domain proteins from *Oryza sativa*. Plant Cell Environ. *31*, 995–1006.

Chaikam, V., and Karlson, D.T. (2010). Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. BMB Rep *43*, 1–8.

Chandrasekar, S., Chartron, J., Jaru-Ampornpan, P., and Shan, S. (2008). Structure of the chloroplast signal recognition particle (SRP) receptor: domain arrangement modulates SRP-receptor interaction. J. Mol. Biol. *375*, 425–436.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chernov, A.A. (2003). Protein crystals and their growth. Journal of Structural Biology *142*, 3–21.

Choe, S., Bennett, M.J., Fujii, G., Curmi, P.M., Kantardjieff, K.A., Collier, R.J., and Eisenberg, D. (1992). The crystal structure of diphtheria toxin. Nature *357*, 216–222.

Choi, H.K., Kim, K.H., Lee, S.H., and Lee, S.J. (2011). Risk factors for recurrence of *Clostridium difficile* infection: effect of vancomycin-resistant enterococci colonization. J. Korean Med. Sci. *26*, 859–864.

Chumbler, N.M., Farrow, M.A., Lapierre, L.A., Franklin, J.L., and Lacy, D.B. (2016a). *Clostridium difficile* Toxins TcdA and TcdB Cause Colonic Tissue Damage by Distinct Mechanisms. Infect. Immun. *84*, 2871–2877.

Chumbler, N.M., Rutherford, S.A., Zhang, Z., Farrow, M.A., Lisher, J.P., Farquhar, E., Giedroc, D.P., Spiller, B.W., Melnyk, R.A., and Lacy, D.B. (2016b). Crystal structure of *Clostridium difficile* toxin A. Nat Microbiol *1*, 15002.

Cladière, L., Hamze, K., Madec, E., Levdikov, V.M., Wilkinson, A.J., Holland, I.B., and Séror, S.J. (2006). The GTPase, CpgA(YloQ), a putative translation factor, is implicated in morphogenesis in *Bacillus subtilis*. Mol. Genet. Genomics *275*, 409–420.

Cordell, S.C., and Löwe, J. (2001). Crystal structure of the bacterial cell division regulator MinD. FEBS Lett. *492*, 160–165.

Cotter, P.D., Ross, R.P., and Hill, C. (2013). Bacteriocins - a viable alternative to antibiotics? Nat. Rev. Microbiol. *11*, 95–105.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., and McPherson, A. (1994). Screening and optimization strategies for macromolecular crystal growth. Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr *50*, 414–423.

Daigle, D.M., and Brown, E.D. (2004). Studies of the interaction of *Escherichia coli* YjeQ with the ribosome *in vitro*. J. Bacteriol. *186*, 1381–1387.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. *32*, W615-619.

Deakin, L.J., Clare, S., Fagan, R.P., Dawson, L.F., Pickard, D.J., West, M.R., Wren, B.W., Fairweather, N.F., Dougan, G., and Lawley, T.D. (2012). The *Clostridium difficile* spo0A gene is a persistence and transmission factor. Infect. Immun. *80*, 2704–2711.

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography, *40,* 82-92.

Delbrück, H., Mueller, U., Perl, D., Schmid, F.X., and Heinemann, U. (2001). Crystal structures of mutant forms of the *Bacillus caldolyticus* cold shock protein differing in thermal stability. J. Mol. Biol. *313*, 359–369.

Delmée, M., Avesani, V., Delferriere, N., and Burtonboy, G. (1990). Characterization of flagella of *Clostridium difficile* and their role in serogrouping reactions. J. Clin. Microbiol. *28*, 2210–2214.

Dembek, M., Barquist, L., Boinett, C.J., Cain, A.K., Mayho, M., Lawley, T.D., Fairweather, N.F., and Fagan, R.P. (2015). High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. MBio *6*, e02383.

Dessau, M.A., and Modis, Y. (2011a). Protein crystallization for X-ray crystallography. J Vis Exp.

Egerer, M., Giesemann, T., Herrmann, C., and Aktories, K. (2009). Autocatalytic processing of *Clostridium difficile* toxin B. Binding of inositol hexakisphosphate. J. Biol. Chem. *284*, 3389–3395.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of *Coot*. Acta Crystallographica Section D Biological Crystallography *66*, 486–501.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr *69*, 1204–1214.

Ewald, P. P. (1913). Zur Theorie der Interferenzen der Röntgenstrahlen in Kristallen. Physikalische Zeitschrift *11*, 465-472.

Eyre, D.W., Walker, A.S., Griffiths, D., Wilcox, M.H., Wyllie, D.H., Dingle, K.E., Crook, D.W., and Peto, T.E.A. (2012a). *Clostridium difficile* mixed infection and reinfection. J. Clin. Microbiol. *50*, 142–144.

Eyre, D.W., Walker, A.S., Wyllie, D., Dingle, K.E., Griffiths, D., Finney, J., O'Connor, L., Vaughan, A., Crook, D.W., Wilcox, M.H., et al. (2012b). Predictors of first recurrence of *Clostridium difficile* infection: implications for initial management. Clin. Infect. Dis. *55 Suppl 2*, S77-87.

Fagan, R.P., and Fairweather, N.F. (2014). Biogenesis and functions of bacterial S-layers. Nature Reviews Microbiology *12*, 211–222.

Farrow, M.A., Chumbler, N.M., Lapierre, L.A., Franklin, J.L., Rutherford, S.A., Goldenring, J.R., and Lacy, D.B. (2013). *Clostridium difficile* toxin B-induced necrosis is mediated by the host epithelial cell NADPH oxidase complex. Proc. Natl. Acad. Sci. U.S.A. *110*, 18674–18679.

Fekety, R., McFarland, L.V., Surawicz, C.M., Greenberg, R.N., Elmer, G.W., and Mulligan, M.E. (1997). Recurrent *Clostridium difficile* diarrhea: characteristics of and risk factors for patients enrolled in a prospective, randomized, double-blinded trial. Clin. Infect. Dis. *24*, 324–333.

Fenton, H.J.H. (1894). LXXIII.—Oxidation of tartaric acid in presence of iron. J. Chem. Soc., Trans. *65*, 899–910.

Finkelstein, A.V., Gutun, A.M., and Badretdinov AYa, null (1993). Why are the same protein folds used to perform different functions? FEBS Lett. *325*, 23–28.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. *45*, D190–D199.

Flynn, R.L., and Zou, L. (2010). Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. Crit. Rev. Biochem. Mol. Biol. *45*, 266–275.

Foadi, J., and Evans, G. (2008). Elucidations on the reciprocal lattice and the Ewald sphere. Eur. J. Phys. *29*, 1059.

Fokine, A., Kostyuchenko, V.A., Efimov, A.V., Kurochkina, L.P., Sykilinda, N.N., Robben, J., Volckaert, G., Hoenger, A., Chipman, P.R., Battisti, A.J., et al. (2005). A three-dimensional cryo-electron microscopy structure of the bacteriophage phiKZ head. J. Mol. Biol. *352*, 117–124.

Fournet, G., and A. Guinier. Small angle scattering of X-rays. (*Translated by Walker, CB and Yudowitch, KL). New York: John Wiley & Sons*. 7-78.

Francis, M.B., Allen, C.A., Shrestha, R., and Sorg, J.A. (2013a). Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. PLoS Pathog. *9*, e1003356.

Francis, M.B., Allen, C.A., and Sorg, J.A. (2013b). Muricholic acids inhibit *Clostridium difficile* spore germination and growth. PLoS ONE *8*, e73653.

Franke, D., and Svergun, D.I. (2009). DAMMIF , a program for rapidab-initioshape determination in small-angle scattering. J Appl Crystallogr *42*, 342–346.

Fuchs, M.R., Pradervand, C., Thominet, V., Schneider, R., Panepucci, E., Grunder, M., Gabadinho, J., Dworkowski, F.S.N., Tomizaki, T., Schneider, J., et al. (2014). D3, the new diffractometer for the macromolecular crystallography beamlines of the Swiss Light Source. J Synchrotron Rad, J Synchrotron Radiat *21*, 340–351.

Garman, E. (2003). 'Cool' crystals: macromolecular cryocrystallography and radiation damage. Current Opinion in Structural Biology *13*, 545–551.

Gasper, R., Scrima, A., and Wittinghofer, A. (2006). Structural insights into HypB, a GTP-binding protein that regulates metal binding. J. Biol. Chem. *281*, 27492–27502.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. *31*, 3784–3788.

Ge, P., Scholl, D., Leiman, P.G., Yu, X., Miller, J.F., and Zhou, Z.H. (2015). Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. Nat. Struct. Mol. Biol. *22*, 377–382.

Gebhart, D., Williams, S.R., Bishop-Lilly, K.A., Govoni, G.R., Willner, K.M., Butani, A., Sozhamannan, S., Martin, D., Fortier, L.-C., and Scholl, D. (2012). Novel high-molecular-weight, R-type bacteriocins of *Clostridium difficile*. J. Bacteriol. *194*, 6240–6247.

Gebhart, D., Lok, S., Clare, S., Tomas, M., Stares, M., Scholl, D., Donskey, C.J., Lawley, T.D., and Govoni, G.R. (2015). A modified R-type bacteriocin specifically targeting *Clostridium difficile* prevents colonization of mice without affecting gut microbiota diversity. MBio *6*.

Genisyuerek, S., Papatheodorou, P., Guttenberg, G., Schubert, R., Benz, R., and Aktories, K. (2011). Structural determinants for membrane insertion, pore formation and translocation of *Clostridium difficile* toxin B. Mol. Microbiol. *79*, 1643–1654.

Grant, T.D., Luft, J.R., Carter, L.G., Matsui, T., Weiss, T.M., Martel, A., and Snell, E.H. (2015). The accurate assessment of small-angle X-ray scattering data. Acta Crystallogr D Biol Crystallogr *71*, 45–56.

Gras, S., Chaumont, V., Fernandez, B., Carpentier, P., Charrier-Savournin, F., Schmitt, S., Pineau, C., Flament, D., Hecker, A., Forterre, P., et al. (2007). Structural insights into a new homodimeric self-activated GTPase family. EMBO Rep. *8*, 569–575.

Graumann, P., and Marahiel, M.A. (1996a). A case of convergent evolution of nucleic acid binding modules. Bioessays *18*, 309–315.

Graumann, P., and Marahiel, M.A. (1996b). Some like it cold: response of microorganisms to cold shock. Arch. Microbiol. *166*, 293–300.

Graumann, P., Schröder, K., Schmid, R., and Marahiel, M.A. (1996). Cold shock stress-induced proteins in *Bacillus subtilis*. J. Bacteriol. *178*, 4611–4619.

Haas, D.J., and Rossmann, M.G. (1970). Crystallographic studies on lactate dehydrogenase at-75 degrees C. Acta Crystallogr B *26*, 998–1004.

Hall, I.C., and O'toole, E. (1935). INTESTINAL FLORA IN NEW-BORN INFANTS: WITH A DESCRIPTION OF A NEW PATHOGENIC ANAEROBE, *BACILLUS DIFFICILIS*. Am J Dis Child *49*, 390–402.

Hatzopoulos, G.N., and Mueller-Dieckmann, J. (2010). Structure of translation initiation factor 1 from *Mycobacterium tuberculosis* and inferred binding to the 30S ribosomal subunit. FEBS Lett. *584*, 1011–1015.

Hermann, J.C., Ghanem, E., Li, Y., Raushel, F.M., Irwin, J.J., and Shoichet, B.K. (2006). Predicting substrates by docking high-energy intermediates to enzyme structures. J. Am. Chem. Soc. *128*, 15882–15891.

Hermann, J.C., Marti-Arbona, R., Fedorov, A.A., Fedorov, E., Almo, S.C., Shoichet, B.K., and Raushel, F.M. (2007). Structure-based activity prediction for an enzyme of unknown function. Nature *448*, 775–779.

Higgins, D., and Dworkin, J. (2012). Recent progress in *Bacillus subtilis* sporulation. FEMS Microbiol. Rev. *36*, 131–148.

Ho, J.G.S., Greco, A., Rupnik, M., and Ng, K.K.-S. (2005). Crystal structure of receptor-binding C-terminal repeats from *Clostridium difficile* toxin A. Proceedings of the National Academy of Sciences *102*, 18373–18378.

Holm, L., and Laakso, L.M. (2016). Dali server update. Nucleic Acids Res. *44*, W351-355.

Hurst, M.R.H., Glare, T.R., and Jackson, T.A. (2004). Cloning *Serratia entomophila* antifeeding genes--a putative defective prophage active against the grass grub Costelytra zealandica. J. Bacteriol. *186*, 5116–5128.

Hutchinson, E.G., and Thornton, J.M. (1993). The Greek key motif: extraction, classification and analysis. Protein Eng. *6*, 233–245.

Huynen, M.A., and Snel, B. (2000). Gene and context: integrative approaches to genome analysis. Adv. Protein Chem. *54*, 345–379.

Huynen, M., Snel, B., Lathe, W., and Bork, P. (2000a). Exploitation of gene context. Curr. Opin. Struct. Biol. *10*, 366–370.

Huynen, M., Snel, B., Lathe, W., and Bork, P. (2000b). Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res. *10*, 1204–1210.

Huynh, K., and Partch, C.L. (2015). Analysis of protein stability and ligand interactions by thermal shift assay. Curr Protoc Protein Sci *79*, 28.9.1-14.

Jacob, F. (1954). [Induced biosynthesis and mode of action of a pyocine, antibiotic produced by *Pseudomonas aeruginosa*. Ann Inst Pasteur (Paris) *86*, 149–160.

Jank, T., and Aktories, K. (2008). Structure and mode of action of clostridial glucosylating toxins: the ABCD model. Trends in Microbiology *16*, 222–229.

Jank, T., Reinert, D.J., Giesemann, T., Schulz, G.E., and Aktories, K. (2005). Change of the donor substrate specificity of *Clostridium difficile* toxin B by site-directed mutagenesis. J. Biol. Chem. *280*, 37833–37838.

Jank, T., Giesemann, T., and Aktories, K. (2007). *Clostridium difficile* glucosyltransferase toxin B-essential amino acids for substrate binding. J. Biol. Chem. *282*, 35222–35231.

Janoir, C., Denève, C., Bouttier, S., Barbut, F., Hoys, S., Caleechum, L., Chapetón-Montes, D., Pereira, F.C., Henriques, A.O., Collignon, A., et al. (2013). Adaptive strategies and pathogenesis of *Clostridium difficile* from *in vivo* transcriptomics. Infect. Immun. *81*, 3757–3769.

Jump, R.L. (2013). *Clostridium difficile* infection in older adults. Aging Health *9*, 403–414.

Jung Lee, W., Lattimer, L.D.N., Stephen, S., Borum, M.L., and Doman, D.B. (2015). Fecal Microbiota Transplantation: A Review of Emerging Indications Beyond Relapsing *Clostridium difficile* Toxin Colitis. Gastroenterol Hepatol (N Y) *11*, 24–32.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2002). Structure of the cell-puncturing device of bacteriophage T4. Nature *415*, 553–557.

Karjalainen, T., Waligora-Dupriet, A.J., Cerquetti, M., Spigaglia, P., Maggioni, A., Mauri, P., and Mastrantonio, P. (2001). Molecular and genomic analysis of genes encoding surface-anchored proteins from *Clostridium difficile*. Infect. Immun. *69*, 3442–3446.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols *10*, 845–858.

Kelly, B.J., and Tebas, P. (2018). Clinical Practice and Infrastructure Review of Fecal Microbiota Transplantation for Clostridium difficile Infection. Chest *153*, 266–277.

Kibbe, W.A. (2007). OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. *35*, W43-46.

Kim, D.J., Jang, J.Y., Yoon, H.-J., and Suh, S.W. (2008a). Crystal structure of YlqF, a circularly permuted GTPase: implications for its GTPase activation in 50 S ribosomal subunit assembly. Proteins *72*, 1363–1370.

Kim, D.J., Jang, J.Y., Yoon, H.-J., and Suh, S.W. (2008b). Crystal structure of YlqF, a circularly permuted GTPase: implications for its GTPase activation in 50 S ribosomal subunit assembly. Proteins *72*, 1363–1370.

Kim, J.W., Lee, K.L., Jeong, J.B., Kim, B.G., Shin, S., Kim, J.S., Jung, H.C., and Song, I.S. (2010). Proton pump inhibitors as a risk factor for recurrence of *Clostridium-difficile*-associated diarrhea. World J. Gastroenterol. *16*, 3573–3577.

Koch, M.H., Vachette, P., and Svergun, D.I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q. Rev. Biophys. *36*, 147–227.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Cryst, J Appl Crystallogr *36*, 1277–1282.

Konc, J., Hodošček, M., Ogrizek, M., Trykowska Konc, J., and Janežič, D. (2013). Structure-based function prediction of uncharacterized protein using binding sites comparison. PLoS Comput. Biol. *9*, e1003341.

Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., Kanamaru, S., van Raaij, M.J., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2003). Three-dimensional structure of bacteriophage T4 baseplate. Nat. Struct. Biol. *10*, 688–693.

Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2005). The tail structure of bacteriophage T4 and its mechanism of contraction. Nat. Struct. Mol. Biol. *12*, 810–813.

Kovacs-Simon, A., Leuzzi, R., Kasendra, M., Minton, N., Titball, R.W., and Michell, S.L. (2014). Lipoprotein CD0873 is a novel adhesin of *Clostridium difficile*. J. Infect. Dis. *210*, 274–284.

Kreimeyer, I., Euler, F., Marckscheffel, A., Tatge, H., Pich, A., Olling, A., Schwarz, J., Just, I., and Gerhard, R. (2011). Autoproteolytic cleavage mediates cytotoxicity of *Clostridium difficile* toxin A. Naunyn Schmiedebergs Arch. Pharmacol. *383*, 253–262.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Kuijper, E.J., van den Berg, R.J., Debast, S., Visser, C.E., Veenendaal, D., Troelstra, A., van der Kooi, T., van den Hof, S., and Notermans, D.W. (2006a). *Clostridium difficile* ribotype 027, toxinotype III, the Netherlands. Emerging Infect. Dis. *12*, 827–830.

Kuijper, E.J., Coignard, B., Tüll, P., ESCMID Study Group for *Clostridium difficile*, EU Member States, and European Centre for Disease Prevention and Control (2006b). Emergence of *Clostridium difficile*-associated disease in North America and Europe. Clin. Microbiol. Infect. *12 Suppl 6*, 2–18.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

LaFrance, M.E., Farrow, M.A., Chandrasekaran, R., Sheng, J., Rubin, D.H., and Lacy, D.B. (2015). Identification of an epithelial cell receptor responsible for *Clostridium difficile* TcdB-induced cytotoxicity. Proc. Natl. Acad. Sci. U.S.A. *112*, 7073–7078.

Laskowski, R.A., Watson, J.D., and Thornton, J.M. (2005). ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res *33*, W89–W93.

Lee, G., Chakraborty, U., Gebhart, D., Govoni, G.R., Zhou, Z.H., and Scholl, D. (2016). F-Type Bacteriocins of *Listeria monocytogenes*: a New Class of Phage Tail-Like Structures Reveals Broad Parallel Coevolution between Tailed Bacteriophages and High-Molecular-Weight Bacteriocins. J. Bacteriol. *198*, 2784–2793.

Lee, J., Jeong, K.-W., Jin, B., Ryu, K.-S., Kim, E.-H., Ahn, J.-H., and Kim, Y. (2013). Structural and dynamic features of cold-shock proteins of *Listeria monocytogenes*, a psychrophilic bacterium. Biochemistry *52*, 2492–2504.

Lee, J.-H., Yeo, W.-S., and Roe, J.-H. (2004). Induction of the sufA operon encoding Fe-S assembly proteins by superoxide generators and hydrogen peroxide: involvement of OxyR, IHF and an unidentified oxidant-responsive factor. Mol. Microbiol. *51*, 1745–1755.

Leiman, P.G., and Shneider, M.M. (2012). Contractile tail machines of bacteriophages. Adv. Exp. Med. Biol. *726*, 93–114.

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2004). Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. Cell *118*, 419–429.

Leiman, P.G., Battisti, A.J., Bowman, V.D., Stummeyer, K., Mühlenhoff, M., Gerardy-Schahn, R., Scholl, D., and Molineux, I.J. (2007). The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. J. Mol. Biol. *371*, 836–849.

Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., Burley, S.K., Almo, S.C., and Mekalanos, J.J. (2009). Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proc. Natl. Acad. Sci. U.S.A. *106*, 4154–4159.

Leiman, P.G., Arisaka, F., van Raaij, M.J., Kostyuchenko, V.A., Aksyuk, A.A., Kanamaru, S., and Rossmann, M.G. (2010). Morphogenesis of the T4 tail and tail fibers. Virol. J. *7*, 355.

Leipe, D.D., Wolf, Y.I., Koonin, E.V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. J. Mol. Biol. *317*, 41–72.

Levdikov, V.M., Blagova, E.V., Brannigan, J.A., Cladière, L., Antson, A.A., Isupov, M.N., Séror, S.J., and Wilkinson, A.J. (2004). The crystal structure of YloQ, a circularly permuted GTPase essential for *Bacillus subtilis* viability. J. Mol. Biol. *340*, 767–782.

Levitt, M., and Chothia, C. (1976). Structural patterns in globular proteins. Nature *261*, 552–558.

Lindsay, M.A. (2005). Finding new drug targets in the 21st century. Drug Discov. Today *10*, 1683–1687.

Liu, Y., Bauer, S.C., and Imlay, J.A. (2011). The YaaA protein of the *Escherichia coli* OxyR regulon lessens hydrogen peroxide toxicity by diminishing the amount of intracellular unincorporated iron. J. Bacteriol. *193*, 2186–2196.

Lübbert, C., Zimmermann, L., Borchert, J., Hörner, B., Mutters, R., and Rodloff, A.C. (2016). Epidemiology and Recurrence Rates of *Clostridium difficile* Infections in Germany: A Secondary Data Analysis. Infect Dis Ther *5*, 545–554.

Manse, J.S., and Baldwin, M.R. (2015). Binding and entry of *Clostridium difficile* toxin B is mediated by multiple domains. FEBS Lett. *589*, 3945–3951.

Mao, F., Dam, P., Chou, J., Olman, V., and Xu, Y. (2009). DOOR: a database for prokaryotic operons. Nucleic Acids Res. *37*, D459-463.

Mao, X., Ma, Q., Zhou, C., Chen, X., Zhang, H., Yang, J., Mao, F., Lai, W., and Xu, Y. (2014). DOOR 2.0: presenting operons and their functions through dynamic and integrated views. Nucleic Acids Res. *42*, D654-659.

Martin, J., and Wilcox, M. (2016). New and emerging therapies for *Clostridium difficile* infection. Curr. Opin. Infect. Dis. *29*, 546–554.

Martin, J.S.H., Monaghan, T.M., and Wilcox, M.H. (2016). *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. Nat Rev Gastroenterol Hepatol *13*, 206–216.

Matsui, H., Sano, Y., Ishihara, H., and Shinomiya, T. (1993). Regulation of pyocin genes in *Pseudomonas aeruginosa* by positive (*prtN*) and negative (*prtR*) regulatory genes. J. Bacteriol. *175*, 1257–1263.

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. *63*, 32–41.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658–674.

McFarland, L.V., Surawicz, C.M., Rubin, M., Fekety, R., Elmer, G.W., and Greenberg, R.N. (1999). Recurrent *Clostridium difficile* disease: epidemiology and clinical characteristics. Infect Control Hosp Epidemiol *20*, 43–50.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404–405.

McPherson, A., and Gavira, J.A. (2013). Introduction to protein crystallization. Acta Crystallogr F Struct Biol Commun *70*, 2–20.

Meents, A., Reime, B., Stuebe, N., Fischer, P., Warmer, M., Goeries, D., Roever, J., Meyer, J., Fischer, J., Burkhardt, A., et al. (2013). Development of an in-vacuum x-ray microscope with cryogenic sample cooling for beamline P11 at PETRA III. In X-Ray Nanoimaging: Instruments and Methods, (International Society for Optics and Photonics), p. 88510K.

Mesyanzhinov, V.V., Leiman, P.G., Kostyuchenko, V.A., Kurochkina, L.P., Miroshnikov, K.A., Sykilinda, N.N., and Shneider, M.M. (2004). Molecular architecture of bacteriophage T4. Biochemistry Mosc. *69*, 1190–1202.

Michel-Briand, Y., and Baysse, C. (2002). The pyocins of *Pseudomonas aeruginosa*. Biochimie *84*, 499–510.

Morimoto, T., Loh, P.C., Hirai, T., Asai, K., Kobayashi, K., Moriya, S., and Ogasawara, N. (2002). Six GTP-binding proteins of the Era/Obg family are essential for cell growth in *Bacillus subtilis*. Microbiology (Reading, Engl.) *148*, 3539–3552.

Mueller, M., Wang, M., and Schulze-Briese, C. (2012a). Optimal fine φ-slicing for single-photon-counting pixel detectors. Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr *68*, 42–56.

Mueller, U., Darowski, N., Fuchs, M.R., Förster, R., Hellmig, M., Paithankar, K.S., Pühringer, S., Steffien, M., Zocher, G., and Weiss, M.S. (2012b). Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. J Synchrotron Rad, J Synchrotron Radiat *19*, 442–449.

Mueller, U., Förster, R., Hellmig, M., Huschmann, F.U., Kastner, A., Malecki, P., Pühringer, S., Röwer, M., Sparta, K., Steffien, M., et al. (2015). The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. Eur. Phys. J. Plus *130*, 141.

Murase, T., Eugenio, L., Schorr, M., Hussack, G., Tanha, J., Kitova, E.N., Klassen, J.S., and Ng, K.K.S. (2014). Structural basis for antibody recognition in the receptor-binding domains of toxins A and B from *Clostridium difficile*. J. Biol. Chem. *289*, 2331–2343.

Mylonas, E., and Svergun, D.I. (2007). Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. Journal of Applied Crystallography *40*, s245–s249.

Na, X., Kim, H., Moyer, M.P., Pothoulakis, C., and LaMont, J.T. (2008). gp96 is a human colonocyte plasma membrane binding protein for *Clostridium difficile* toxin A. Infect. Immun. *76*, 2862–2871.

Nassi, S., Collier, R.J., and Finkelstein, A. (2002). PA63 channel of anthrax toxin: an extended beta-barrel. Biochemistry *41*, 1445–1450.

Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443–453.

Nováček, J., Šiborová, M., Benešík, M., Pantůček, R., Doškař, J., and Plevka, P. (2016). Structure and genome release of Twort-like Myoviridae phage with a double-layered baseplate. Proc. Natl. Acad. Sci. U.S.A. *113*, 9351–9356.

Odumosu, O., Nicholas, D., Yano, H., and Langridge, W. (2010). AB Toxins: A Paradigm Switch from Deadly to Desirable. Toxins (Basel) *2*, 1612–1645.

Oliveira, S.H., Ferraz, F.A., Honorato, R.V., Xavier-Neto, J., Sobreira, T.J., and de Oliveira, P.S. (2014). KVFinder: steered identification of protein cavities as a PyMOL plugin. BMC Bioinformatics *15*, 197.

Olling, A., Goy, S., Hoffmann, F., Tatge, H., Just, I., and Gerhard, R. (2011). The repetitive oligopeptide sequences modulate cytopathic potency but are not crucial for cellular uptake of *Clostridium difficile* toxin A. PLoS ONE *6*, e17623.

Olling, A., Seehase, S., Minton, N.P., Tatge, H., Schröter, S., Kohlscheen, S., Pich, A., Just, I., and Gerhard, R. (2012). Release of TcdA and TcdB from *Clostridium difficile* cdi 630 is not affected by functional inactivation of the tcdE gene. Microb. Pathog. *52*, 92–100.

Orth, P., Xiao, L., Hernandez, L.D., Reichert, P., Sheth, P.R., Beaumont, M., Yang, X., Murgolo, N., Ermakov, G., DiNunzio, E., et al. (2014). Mechanism of action and epitopes of *Clostridium difficile* toxin B-neutralizing antibody bezlotoxumab revealed by X-ray crystallography. J. Biol. Chem. *289*, 18008–18021.

Pal, D., and Eisenberg, D. (2005). Inference of protein function from protein structure. Structure *13*, 121–130.

Papatheodorou, P., Zamboglou, C., Genisyuerek, S., Guttenberg, G., and Aktories, K. (2010). Clostridial glucosylating toxins enter cells via clathrin-mediated endocytosis. PLoS ONE *5*, e10673.

Paredes-Sabja, D., Shen, A., and Sorg, J.A. (2014). *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol. *22*, 406–416.

Pei, J., and Grishin, N.V. (2014). PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol. Biol. *1079*, 263–271.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J Synchrotron Rad, J Synchrotron Radiat *20*, 660–664.

Petoukhov, M.V., and Svergun, D.I. (2013). Applications of small-angle X-ray scattering to biomacromolecular solutions. Int. J. Biochem. Cell Biol. *45*, 429–437.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605–1612.

Pfeifer, G., Schirmer, J., Leemhuis, J., Busch, C., Meyer, D.K., Aktories, K., and Barth, H. (2003). Cellular uptake of *Clostridium difficile* toxin B. Translocation of the N-terminal catalytic domain into the cytosol of eukaryotic cells. J. Biol. Chem. *278*, 44535–44541.

Porod, G., Small angle X-ray Scattering (1982) (eds. Glatter & Kratky). Academic Press, London.

Powell, H.R. (2017). X-ray data processing. Biosci Rep *37*.

Pruitt, R.N., Chagot, B., Cover, M., Chazin, W.J., Spiller, B., and Lacy, D.B. (2009). Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in *Clostridium difficile* toxin A. J. Biol. Chem. *284*, 21934–21940.

Pruitt, R.N., Chambers, M.G., Ng, K.K.-S., Ohi, M.D., and Lacy, D.B. (2010). Structural organization of the functional domains of *Clostridium difficile* toxins A and B. Proc. Natl. Acad. Sci. U.S.A. *107*, 13467–13472.

Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., and Mekalanos, J.J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. PNAS *104*, 15508–15513.

Puri, A.W., Lupardus, P.J., Deu, E., Albrow, V.E., Garcia, K.C., Bogyo, M., and Shen, A. (2010). Rational design of inhibitors and activity-based probes targeting *Clostridium difficile* virulence factor TcdB. Chem. Biol. *17*, 1201–1211.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q. Rev. Biophys. *40*, 191–285.

Putnam, D.K., Lowe, E.W., and Meiler, J. (2013). Reconstruction of SAXS Profiles from Protein Structures. Comput Struct Biotechnol J *8*, e201308006.

Qa'Dan, M., Spyres, L.M., and Ballard, J.D. (2000). pH-induced conformational changes in *Clostridium difficile* toxin B. Infect. Immun. *68*, 2470–2474.

Ramakrishnan, C., and Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. Biophys. J. *5*, 909–933.

Reinert, D.J., Jank, T., Aktories, K., and Schulz, G.E. (2005). Structural basis for the function of *Clostridium difficile* toxin B. J. Mol. Biol. *351*, 973–981.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320–W324.

Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., and Leiman, P.G. (2004). The bacteriophage T4 DNA injection machine. Curr. Opin. Struct. Biol. *14*, 171–180.

Rupp, B. (2009). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology New York: Garland Science.

Rupp, B. (2015). Origin and use of crystallization phase diagrams. Acta Cryst F, Acta Cryst Sect F, Acta Crystallogr F, Acta Crystallogr Sect F, Acta Cryst F Struct Biol Cryst Commun, Acta Cryst Sect F Struct Biol Cryst Commun, Acta Crystallogr Sect F Struct Biol Cryst Commun *71*, 247–260.

Sangster, W., Hegarty, J.P., and Stewart, D.B. (2015). Phage tail-like particles kill *Clostridium difficile* and represent an alternative to conventional antibiotics. Surgery *157*, 96–103.

Sára, M., and Sleytr, U.B. (2000). S-Layer proteins. J. Bacteriol. *182*, 859–868.

Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. *15*, 430–434.

Schindelin, H., Marahiel, M.A., and Heinemann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. Nature *364*, 164–168.

Schirmer, J., and Aktories, K. (2004). Large clostridial cytotoxins: cellular biology of Rho/Ras-glucosylating toxins. Biochim. Biophys. Acta *1673*, 66–74.

Schorch, B., Song, S., van Diemen, F.R., Bock, H.H., May, P., Herz, J., Brummelkamp, T.R., Papatheodorou, P., and Aktories, K. (2014). LRP1 is a receptor for *Clostridium perfringens* TpeL toxin indicating a two-receptor model of clostridial glycosylating toxins. Proc. Natl. Acad. Sci. U.S.A. *111*, 6431–6436.

Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeño-Tárraga, A.M., Wang, H., et al. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat. Genet. *38*, 779–786.

Serdyuk, I.N., Tsalkova, T.N., Svergun, D.I., and Izotova, T.D. (1987). Determination of radii of gyration of particles by small-angle neutron scattering: calculation of the effect of aggregates. J. Mol. Biol. *194*, 126–128.

Sette, M., van Tilborg, P., Spurio, R., Kaptein, R., Paci, M., Gualerzi, C.O., and Boelens, R. (1997). The structure of the translational initiation factor IF1 from *E.coli* contains an oligomer-binding motif. EMBO J. *16*, 1436–1443.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr., A, Found. Crystallogr. *64*, 112–122.

Shen, A. (2012). *Clostridium difficile* toxins: mediators of inflammation. J Innate Immun *4*, 149–158.

Shen, A., Lupardus, P.J., Gersch, M.M., Puri, A.W., Albrow, V.E., Garcia, K.C., and Bogyo, M. (2011). Defining an allosteric circuit in the cysteine protease domain of *Clostridium difficile* toxins. Nat. Struct. Mol. Biol. *18*, 364–371.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Slater, L.H., Hett, E.C., Mark, K., Chumbler, N.M., Patel, D., Lacy, D.B., Collier, R.J., and Hung, D.T. (2013). Identification of novel host-targeted compounds that protect from anthrax lethal toxin-induced cell death. ACS Chem. Biol. *8*, 812–822.

Smilgies, D.-M., and Folta-Stogniew, E. (2015a). Molecular weight-gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data. J Appl Crystallogr *48*, 1604–1606.

Smilgies, D.-M., and Folta-Stogniew, E. (2015b). Molecular weight–gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data. J Appl Crystallogr *48*, 1604–1606.

Smith, S.M.E., Min, J., Ganesh, T., Diebold, B., Kawahara, T., Zhu, Y., McCoy, J., Sun, A., Snyder, J.P., Fu, H., et al. (2012). Ebselen and congeners inhibit NADPH oxidase 2-dependent superoxide generation by interrupting the binding of regulatory subunits. Chem. Biol. *19*, 752–763.

Smits, W.K., Lyras, D., Lacy, D.B., Wilcox, M.H., and Kuijper, E.J. (2016). *Clostridium difficile* infection. Nat Rev Dis Primers *2*, 16020.

Sorg, J.A., and Sonenshein, A.L. (2008). Bile salts and glycine as cogerminants for *Clostridium difficile* spores. J. Bacteriol. *190*, 2505–2512.

Sorg, J.A., and Sonenshein, A.L. (2009). Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. J. Bacteriol. *191*, 1115–1117.

Steiner, E., Dago, A.E., Young, D.I., Heap, J.T., Minton, N.P., Hoch, J.A., and Young, M. (2011). Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. Mol Microbiol *80*, 641–654.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Sudhamsu, J., Lee, G.I., Klessig, D.F., and Crane, B.R. (2008). The structure of YqeH. An AtNOS1/AtNOA1 ortholog that couples GTP hydrolysis to molecular recognition. J. Biol. Chem. *283*, 32968–32976.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Cryst, J Appl Crystallogr *25*, 495–503.

Svergun, D.I. (2010). Small-angle X-ray and neutron scattering as a tool for structural systems biology. Biol. Chem. *391*, 737–743.

Svergun, D.I., and Koch, M.H.J. (2003). Macromolecules in solution.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J Appl Cryst, J Appl Crystallogr *28*, 768–773.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362–D368.

Tao, L., Zhang, J., Meraner, P., Tovaglieri, A., Wu, X., Gerhard, R., Zhang, X., Stallcup, W.B., Miao, J., He, X., et al. (2016). Frizzled proteins are colonic epithelial receptors for *C. difficile* toxin B. Nature *538*, 350–355.

Tasteyre, A., Barc, M.C., Collignon, A., Boureau, H., and Karjalainen, T. (2001). Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. Infect. Immun. *69*, 7937–7940.

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.-W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr *65*, 582–601.

Theobald, D.L., Mitton-Fry, R.M., and Wuttke, D.S. (2003). Nucleic acid recognition by OB-fold proteins. Annu Rev Biophys Biomol Struct *32*, 115–133.

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. *45*, D158–D169.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem *31*, 455–461.

Vaishnavi, C. (2015). Fidaxomicin--the new drug for *Clostridium difficile* infection. Indian J. Med. Res. *141*, 398–407.

Vary, P.S. (1994). Prime time for *Bacillus megaterium*. Microbiology (Reading, Engl.) *140 (Pt 5)*, 1001–1013.

Waligora, A.J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., and Karjalainen, T. (2001). Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. Infect. Immun. *69*, 2144–2153.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. *1*, 945–951.

Walter, M., Fiedler, C., Grassl, R., Biebl, M., Rachel, R., Hermo-Parrado, X.L., Llamas-Saiz, A.L., Seckler, R., Miller, S., and van Raaij, M.J. (2008). Structure of the receptor-binding protein of bacteriophage det7: a podoviral tail spike in a myovirus. J. Virol. *82*, 2265–2273.

Wang, J., and London, E. (2009). The membrane topography of the diphtheria toxin T domain linked to the a chain reveals a transient transmembrane hairpin and potential translocation mechanisms. Biochemistry *48*, 10446–10456.

Wang, J., Brackmann, M., Castaño-Díez, D., Kudryashev, M., Goldie, K.N., Maier, T., Stahlberg, H., and Basler, M. (2017). Cryo-EM structure of the extended type VI secretion system sheath-tube complex. Nat Microbiol *2*, 1507–1512.

Watson, J.D., Laskowski, R.A., and Thornton, J.M. (2005). Predicting protein function from sequence and structural data. Curr. Opin. Struct. Biol. *15*, 275–284.

Weiss, T.M. (2017). Small Angle Scattering: Historical Perspective and Future Outlook. Adv. Exp. Med. Biol. *1009*, 1–10.

Wheeldon, L.J., Worthington, T., and Lambert, P.A. (2011). Histidine acts as a co-germinant with glycine and taurocholate for *Clostridium difficile* spores. J. Appl. Microbiol. *110*, 987–994.

Wilcox, M., Dorr, M.-B., and Pedley, A. (2017a). Bezlotoxumab and Recurrent *Clostridium difficile* Infection. N. Engl. J. Med. *376*, 1594–1596.

Wilcox, M.H., Gerding, D.N., Poxton, I.R., Kelly, C., Nathan, R., Birch, T., Cornely, O.A., Rahav, G., Bouza, E., Lee, C., et al. (2017b). Bezlotoxumab for Prevention of Recurrent *Clostridium difficile* Infection. N. Engl. J. Med. *376*, 305–317.

Wild, D., and Saqi, M. (2004). Structural Proteomics: Inferring Function from Protein Structure. Current Proteomics *1*, 59–65.

William Studier, F., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). [6] Use of T7 RNA polymerase to direct expression of cloned genes. In Methods in Enzymology, (Academic Press), pp. 60–89.

Williamson, J.R. (2000). Induced fit in RNA-protein recognition. Nat. Struct. Biol. *7*, 834–837.

Wilson, K.H. (1993). The Microecology of *Clostridium difficile*. Clin Infect Dis *16*, S214–S218.

Wriggers, W. (2012). Conventions and workflows for using Situs. Acta Crystallogr D Biol Crystallogr *68*, 344–351.

Yang, G., Dowling, A.J., Gerike, U., ffrench-Constant, R.H., and Waterfield, N.R. (2006). *Photorhabdus* virulence cassettes confer injectable insecticidal activity against the wax moth. J. Bacteriol. *188*, 2254–2261.

Yang, G., Zhou, B., Wang, J., He, X., Sun, X., Nie, W., Tzipori, S., and Feng, H. (2008). Expression of recombinant *Clostridium difficile* toxin A and B in Bacillus megaterium. BMC Microbiol. *8*, 192.

Yang, Z., Ramsey, J., Hamza, T., Zhang, Y., Li, S., Yfantis, H.G., Lee, D., Hernandez, L.D., Seghezzi, W., Furneisen, J.M., et al. (2015). Mechanisms of protection against *Clostridium difficile* infection by the monoclonal antitoxin antibodies actoxumab and bezlotoxumab. Infect. Immun. *83*, 822–831.

Yap, M.L., Mio, K., Leiman, P.G., Kanamaru, S., and Arisaka, F. (2010). The baseplate wedges of bacteriophage T4 spontaneously assemble into hubless baseplate-like structure in vitro. J. Mol. Biol. *395*, 349–360.

Yuan, P., Zhang, H., Cai, C., Zhu, S., Zhou, Y., Yang, X., He, R., Li, C., Guo, S., Li, S., et al. (2015). Chondroitin sulfate proteoglycan 4 functions as the cellular receptor for *Clostridium difficile* toxin B. Cell Res. *25*, 157–168.

Zhang, C., and Kim, S.-H. (2003). Overview of structural genomics: from structure to function. Curr Opin Chem Biol *7*, 28–32.

Zhang, H., and Schumacher, M.A. (2017). Structures of partition protein ParA with nonspecific DNA and ParB effector reveal molecular insights into principles governing Walker-box DNA segregation. Genes Dev. *31*, 481–492.

Zhang, F., Hu, M., Tian, G., Zhang, P., Finley, D., Jeffrey, P.D., and Shi, Y. (2009). Structural insights into the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. Mol. Cell *34*, 473–484.

Zhang, Y., Shi, L., Li, S., Yang, Z., Standley, C., Yang, Z., ZhuGe, R., Savidge, T., Wang, X., and Feng, H. (2013). A segment of 97 amino acids within the translocation domain of *Clostridium difficile* toxin B is essential for toxicity. PLoS ONE *8*, e58634.

Zhang, Z., Park, M., Tam, J., Auger, A., Beilhartz, G.L., Lacy, D.B., and Melnyk, R.A. (2014). Translocation domain mutations affecting cellular toxicity identify the *Clostridium difficile* toxin B pore. Proc. Natl. Acad. Sci. U.S.A. *111*, 3721–3726.

Zhao, S., Kumar, R., Sakai, A., Vetting, M.W., Wood, B.M., Brown, S., Bonanno, J.B., Hillerich, B.S., Seidel, R.D., Babbitt, P.C., et al. (2013). Discovery of new enzymes and metabolic pathways by using structure and genome context. Nature *502*, 698–702.

Zheng, M., Aslund, F., and Storz, G. (1998). Activation of the OxyR transcription factor by reversible disulfide bond formation. Science *279*, 1718–1721.

Zheng, W., Wang, F., Taylor, N.M.I., Guerrero-Ferreira, R.C., Leiman, P.G., and Egelman, E.H. (2017). Refined Cryo-EM Structure of the T4 Tail Tube: Exploring the Lowest Dose Limit. Structure *25*, 1436–1441.e2.

11. Supplement

Table 31: Overview of oligonucleotides used for SLIC cloning of uncharacterized proteins in this study.

Oligonucleotide	Sequence 5' → 3'
p10$-10670fw-Nde	GTTTCAGGGCCCGCATATGCAGGACTACAAAAAAACA
p10$-10670rv-Xho	TTAGCAGCCGGATCCCTCGAGTTAACGATGACAGC
pMal-10670-fwNde	CTTTATTTTCAGGGCCATATGCAGGACTACAAAAAAAAC
pMal-10670-rvXho	CAAGCTTGAATTCCTCGAGTTAACGATGACAGCT
pET19-10670fwNde	GAATCTTTATTTTCAGGGCCATATGCAGGACTACAAAAAAAAC
pET19-10670rvXho pETSumo-1067rvXho	GCAGCCGGATCCCTCGAGTTAACGATGACAGCT
Sumo-10670-fwNde	CGAACAGATCGGTGGTCATATGCAGGACTACAAAAAAAAC
p10$-12190-fwNde	GTTTCAGGGCCCGCATATGATTAGCAAACGTGTGGGTAT
p10$-12190-rvXho pET19-12190rvXho pETSumo-12190rvXho	AGCCGGATCCCTCGAGTTAAACTTCCATCAGATA
pET19-12190fwNde	TCTTTATTTTCAGGGCCATATGATTAGCAAACGTGTGGG
pMal-12190-fwNde	CTTTATTTTCAGGGCCATATGATTAGCAAACGTGTGG
pMal-12190-rvXho	CAAGCTTGAATTCCTCGAGTTAAACTTCCATCAGATAAT
Sumo-12190-fwNde	CGAACAGATCGGTGGTCATATGATTAGCAAACGTGT
p10$-18230-fwNde	GTTTCAGGGCCCGCATATGATCATTATGCTGAGCCC
p10$-18230-rvXho pET19-18230rvXho pETSumo-18230rvXho	TTAGCAGCCGGATCCCTCGAGTTACTCTTTCACAAAGATG
pET19-18230fwNde	GAATCTTTATTTTCAGGGCCATATGATCATTATGCTGAGCCC
pMal-18230-fwNde	CTTTATTTTCAGGGCCATATGATCATTATGCTGAGCC
pMal-18230-rvXho	CAAGCTTGAATTCCTCGAGTTACTCTTTCACAAAGATG
Sumo-18230-fwNde	GAACAGATCGGTGGTCATATGATCATTATGCTGAGCCC
p10$-25890-fwNde	GTTTCAGGGCCCGCATATGGCAATTAGCATGACCG
p10$-25890-rvXho pET19-25890rvXho pETSumo-25890rvXho	TTAGCAGCCGGATCCCTCGAGTTATTCGATGTTCTGGATC
pET19-25890fwNde	GAATCTTTATTTTCAGGGCCATATGGCAATTAGCATGACCG
pMal-25890-fwNde	CTTTATTTTCAGGGCCATATGGCAATTAGCATGACC
pMal-25890-rvXho	CAAGCTTGAATTCCTCGAGTTATTCGATGTTCTGGATC
Sumo-25890-fwNde	CGAACAGATCGGTGGTCATATGGCAATTAGCATGACAGG
p10$-27520-fwNde	TTCAGGGCCCGCATATGAAAGGCAACATCGTGCAG
p10$-27520-rvXho Sumo-27520-rvXho	TTAGCAGCCGGATCCCTCGAGTTAGTTCAGGTTACCGG
pET19-27520fwNde	GAATCTTTATTTTCAGGGCCATATGAAGGGCAACATCGTGCAG
pET19-27520rvXho	CAGCCGGATCCCTCGAGTTAGTTCAGGTTACC
pMal-27520-fwNde	CTTTATTTTCAGGGCCATATGAAGGGAAACATCGTGCAG
pMal-27520-rvXho	CAAGCTTGAATTCCTCGAGTTAGTTCAGGTTACCGGCT
Sumo-27520-fwNde	CGAACAGATCGGTGGTCATATGAAAGGCAACATCGTG

Table 32: Overview of oligonucleotides used for traditional cloning of uncharacterized proteins in this study.

Oligonucleotide	Sequence 5’ → 3’
1067-Nde-fwd	CGCGCCCATATGCAAGATTATAAAAAAAAT
1067-Xho-rev	CTCGAGCTATCTGTGGCAACTTG
1823-Nde-fwd	CGCGCCCATATGATTATAATGTTATCACCAGC
1823-Xho-rev	CGCGCCCTCGAGCTACTCTTTTACAAATATG
2589-Nde-fwd	CCATATGGCAATTAGCATGACCGG
2589-Xho-rev	GCCCTCGAGTTATTCTATGTTCTGGATC
2752-Nde-fwd	GCCCATATGAAAGGCAACATCGTGC
2752-Xho-rev	CGCCTCGAGTTAGTTCAGGTTACC

Table 33: Overview of oligonucleotides used for mutation in this study.

Oligonucleotide	Sequence 5’ → 3’
2589QC-fw1	CGGACATTGTGAAAATCGAAGCGGCAGAAGAGGACGAAGATCTGC
2589QC-rev1	GCAGATCTTCGTCCTCTTCTGCCGCTTCGATTTTCACAATGTCCG
2589QC-fw2	GAAAATCGAAGCGGCAGCGGAGGACGAAGATCTGCTGTG
2589QC-rev2	CACAGCAGATCTTCGTCCTCCGCTGCCGCTTCGATTTTC
QCp10$-1067-1-fw	GGCCCGCATATGCAGGACTACAAAAAAAACAAACGTCGC
QCp10$-1067-1-rv	GCGACGTTTGTTTTTTTTGTAGTCCTGCATATGCGGGCC

Table 34: Overview of oligonucleotides used for SLIC cloning of diffocin proteins during DPF campaign in this study

Oligonucleotide	Sequence 5’ → 3’
pOPIN-His-MBP-CD1363_fwd	GAAGTTCTGTTTCAGGGTCCCATGGCTATAGGATTACCAAGTATCAACA TATCATTTAAG
pOPIN-His-MBP-CD1363_rev	TAAACTGGTCTAGAAAGCTTATATTTCTATTGATAAATCTATATCTTCCAT AGCATCAAGTACTTTTATTTTTG
pOPIN-His-MBP-CD1364_fwd	GAAGTTCTGTTTCAGGGTCCCATGGCAAATATGGAAGCTAGAAATGTAA TGAGT
pOPIN-His-MBP-CD1364_rev	TAAACTGGTCTAGAAAGCTTAAATTATATCAAGAAAATCATACTCAGTAA ATGTGAAAGGTGC
pOPIN-His-MBP-CD1365_fwd	GAAGTTCTGTTTCAGGGTCCCATGAGTGAAAATGGATTATTAAAAAATAT AAACATAGTAGATTTACTTTT
pOPIN-His-MBP-CD1365_rev	TAAACTGGTCTAGAAAGCTTAGTTTTTTACCTCTTCTATCACTGCATTTTT ACCATAAC
pOPIN-His-MBP-CD1366_fwd	GAAGTTCTGTTTCAGGGTCCCATGGGAAATGTGAGAGAAGAAGGTATAA ATATGTAC
pOPIN-His-MBP-CD1366_rev	TAAACTGGTCTAGAAAGCTTATAAGACACCCCCTAACTGTGCC
pOPIN-His-MBP-CD1367_fwd	GAAGTTCTGTTTCAGGGTCCCATGGAAATGTGGCTTAGACAAGCTGAA
pOPIN-His-MBP-CD1367_rev	TAAACTGGTCTAGAAAGCTTATTGTTTTGGAGTTACTTTTGGTATCTTTAT TCTTTTATACTC
pOPIN-His-MBP-CD1368_fwd	GAAGTTCTGTTTCAGGGTCCCATGATAATTAATAGGTCTAAAGATTCAAG CAGTAATGAAATTAGTT
pOPIN-His-MBP-CD1368_rev	TAAACTGGTCTAGAAAGCTTATAATACTCTCCTAGCTGTGTTAAAATCTC TTCTACT

pOPIN-His-MBP-CD1369_fwd	GAAGTTCTGTTTCAGGGTCCCATGGCTAATCCAATAAATGAATTTATAGGAATAATAAGAGAAGAAG
pOPIN-His-MBP-CD1369_rev	TAAACTGGTCTAGAAAGCTTATAAGCTCACCAACTTACTTACAACAGCA
pOPIN-His-MBP-CD1370_fwd	GAAGTTCTGTTTCAGGGTCCCATGAGTACAATATTTCCTTTTATAGGTGTCCCAGA
pOPIN-His-MBP-CD1370_rev	TAAACTGGTCTAGAAAGCTTATACATTTATTTCCACCTCCCCATAGATAGTGG
pOPIN-His-MBP-CD1371_fwd	GAAGTTCTGTTTCAGGGTCCCATGTATAGTGACCAGACATATGAAGTAATAAAAAATAGAACTCT
pOPIN-His-MBP-CD1371_rev	TAAACTGGTCTAGAAAGCTTATTGATTTTCCACCTCACTAAAATTAACAGTTGTTAC
pOPIN-His-MBP-CD1372_fwd	GAAGTTCTGTTTCAGGGTCCCATGAAGCTAATTGATAAACTACCATCATTTGATAGAAATTACA
pOPIN-His-MBP-CD1372_rev	TAAACTGGTCTAGAAAGCTTATACCACCTCCGATTTAGGATAAACTACTAC
pOPIN-His-MBP-CD1373_fwd	GAAGTTCTGTTTCAGGGTCCCTTGGAAGAAAAATTTTATATAATATTAACCAAAATTGGTAGAGAA
pOPIN-His-MBP-CD1373_rev	TAAACTGGTCTAGAAAGCTTATCTTATTTCATACGAATACTTATTTAGAGTTATATCTTTTTGTTC
pOPIN-His-MBP-CD1374_85-536_fwd	GAAGTTCTGTTTCAGGGTCCCTATGTATCTTGTGGTTCAAAACACAGTGTAGC
pOPIN-His-MBP-CD1374_85-536_rev	TAAACTGGTCTAGAAAGCTTAACCCCACCCATATAATTCCTTTGAATGATTTAATG
pOPIN-His-MBP-CD1374_911-1420_fwd	GAAGTTCTGTTTCAGGGTCCCAGTATAGATGGTAACACAACATTTGCAATTAGAAATGA
pOPIN-His-MBP-CD1374_911-1420_rev	TAAACTGGTCTAGAAAGCTTATGTAGTTACTACTAATGTATAGTTTTTACCACAATGTAC
pOPIN-His-MBP-CD1374fl_fwd	GAAGTTCTGTTTCAGGGTCCCATGAAGCAAAATAAACTTTTACAGCGTGGTG
pOPIN-His-MBP-CD1374fl_rev	TAAACTGGTCTAGAAAGCTTAAGCTTGTGGCTCGTTGTTTTCAGC

Table 35: Overview of oligonucleotides used for SLIC cloning of TcdB constructs during DPF campagne in this study.

Oligonucleotide	Sequence 5’ → 3’
pOPIN-His-MBP-NIS1000-1300_fwd	GAAGTTCTGTTTCAGGGTCCCACTATTACAGATGCAGCCAAAGTTGTTGA
pOPIN-His-MBP-NIS1000-1300_rev	TAAACTGGTCTAGAAAGCTTATATAAAACTTCTAGTATTACTATCTAAATTTATTCTTATATTAGTATCTTCATATCTTGGT
pOPIN-His-MBP-NIS1300-1600_fwd	GAAGTTCTGTTTCAGGGTCCCGTTCCAATAATAACTACAGAATATATAAGAGAAAAATTATCATATTCTTTCTATGG
pOPIN-His-MBP-NIS1300-1600_rev	TAAACTGGTCTAGAAAGCTTATTGTAAGAAATTAACGAAAATACTTTTTATATTCATACTTTCTAAAAAGC
pOPIN-His-MBP-NIS1400-1700_fwd	GAAGTTCTGTTTCAGGGTCCCGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAATGC
pOPIN-His-MBP-NIS1400-1700_rev	TAAACTGGTCTAGAAAGCTTATGTATAAATATTTGGTGAAATTACAACTTTATTAACACAACTGT
pOPIN-His-SUMO-NIS1000-1300_fwd	GCGAACAGATCGGTGGTACTATTACAGATGCAGCCAAAGTTGTTGA

pOPIN-His-SUMO-NIS1000-1300_rev	TAAACTGGTCTAGAAAGCTTATATAAAACTTCTAGTATTACTATCTAAATTTATTCTTATATTAGTATCTTCATATCTTGGT
pOPIN-His-SUMO-NIS1300-1600_fwd	GCGAACAGATCGGTGGTGTTCCAATAATAACTACAGAATATATAAGAGAAAAATTATCATATTCTTTCTATGG
pOPIN-His-SUMO-NIS1300-1600_rev	TAAACTGGTCTAGAAAGCTTATTGTAAGAAATTAACGAAAATACTTTTTATATTCATACTTTCTAAAAAGC
pOPIN-His-SUMO-NIS1400-1700_fwd	GCGAACAGATCGGTGGTGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAATGC
pOPIN-His-SUMO-NIS1400-1700_rev	TAAACTGGTCTAGAAAGCTTATGTATAAATATTTGGTGAAATTACAACTTTATTAACACAACTGT
pOPIN-NHis-3C-NIS1000-1300_fwd	GAAGTTCTGTTTCAGGGTCCCACTATTACAGATGCAGCCAAAGTTGTTGA
pOPIN-NHis-3C-NIS1000-1300_rev	CAAACTGGTCTAGAAAGCTTATATAAAACTTCTAGTATTACTATCTAAATTTATTCTTATATTAGTATCTTCATATCTTGGT
pOPIN-NHis-3C-NIS1300-1600_fwd	GAAGTTCTGTTTCAGGGTCCCGTTCCAATAATAACTACAGAATATATAAGAGAAAAATTATCATATTCTTTCTATGG
pOPIN-NHis-3C-NIS1300-1600_rev	CAAACTGGTCTAGAAAGCTTATTGTAAGAAATTAACGAAAATACTTTTTATATTCATACTTTCTAAAAAGC
pOPIN-NHis-3C-NIS1400-1700_fwd	GAAGTTCTGTTTCAGGGTCCCGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAATGC
pOPIN-NHis-3C-NIS1400-1700_rev	CAAACTGGTCTAGAAAGCTTATGTATAAATATTTGGTGAAATTACAACTTTATTAACACAACTGT
pOPIN-NHis-3C-NIS543-2366_fwd	GAAGTTCTGTTTCAGGGTCCCCTTGGTGAAGATGATAATCTTGATTTTTCTCAAAATATAGTAGT
pOPIN-NHis-3C-NIS543-2366_rev	CAAACTGGTCTAGAAAGCTTATTCACTAATCACTAATTGAGCTGTATCAGGATCAAAAT
pOPIN-CHis-NIS1400-1834_fwd	CAATCAAAGGAGATATACCATGGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAA
pOPIN-CHis-NIS1400-1834_rev	GATGGTGATGGTGGTGCTCGAGTCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-CHis-NIS800-1400_fwd	CAATCAAAGGAGATATACCATGTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-CHis-NIS800-1400_rev	GATGGTGATGGTGGTGCTCGAGATTACTTCCATTTACCTCACCAGAAAAATTAATCTC
pOPIN-CHis-NIS800-1834_fwd	CAATCAAAGGAGATATACCATGTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-CHis-NIS800-1834_rev	GATGGTGATGGTGGTGCTCGAGTCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-His-MBP-NIS1100-1400_fwd	GAAGTTCTGTTTCAGGGTCCCTCAGCAGGTATACCAAGCTTAGTAAACAATG
pOPIN-His-MBP-NIS1100-1400_rev	TAAACTGGTCTAGAAAGCTTAATTACTTCCATTTACCTCACCAGAAAAATTAATCTC
pOPIN-His-MBP-NIS1200-1500_fwd	GAAGTTCTGTTTCAGGGTCCCTTATCTATATATGACGTATTGGAAGTACAAAAAGAAGAACTTG
pOPIN-His-MBP-NIS1200-1500_rev	TAAACTGGTCTAGAAAGCTTACATATAAACCTTACTTATAAGAACTACATCAGGTAATTCA
pOPIN-His-MBP-NIS1400-1834_fwd	GAAGTTCTGTTTCAGGGTCCCGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAA
pOPIN-His-MBP-NIS1400-1834_rev	TAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-His-MBP-NIS1500-1834_fwd	GAAGTTCTGTTTCAGGGTCCCGATGATAGTAAGCCTTCATTTGGATATTATAGTAATAATTTGAAAGAT

pOPIN-His-MBP-NIS1500-1834_rev	TAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-His-MBP-NIS800-1100_fwd	GAAGTTCTGTTTCAGGGTCCCTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-His-MBP-NIS800-1100_rev	TAAACTGGTCTAGAAAGCTTAAATTCCTGCTAAAGGAACTAAAAGTATACTAAATCCA
pOPIN-His-MBP-NIS800-1400_fwd	GAAGTTCTGTTTCAGGGTCCCTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-His-MBP-NIS800-1400_rev	TAAACTGGTCTAGAAAGCTTAATTACTTCCATTTACCTCACCAGAAAAATTAATCTC
pOPIN-His-MBP-NIS800-1834_fwd	GAAGTTCTGTTTCAGGGTCCCTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-His-MBP-NIS800-1834_rev	TAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-His-MBP-NIS900-1200_fwd	GAAGTTCTGTTTCAGGGTCCCTTTATTAATAAAGAAACTGGAGAATCTATATTTGTAGAAACTGAAAAAAC
pOPIN-His-MBP-NIS900-1200_rev	TAAACTGGTCTAGAAAGCTTAGTGTGGCTCTCTATATGTTATTGATGGTGCT
pOPIN-His-SUMO-NIS1100-1400_fwd	GCGAACAGATCGGTGGTTCAGCAGGTATACCAAGCTTAGTAAACAATG
pOPIN-His-SUMO-NIS1100-1400_rev	TAAACTGGTCTAGAAAGCTTAATTACTTCCATTTACCTCACCAGAAAAATTAATCTC
pOPIN-His-SUMO-NIS1200-1500_fwd	GCGAACAGATCGGTGGTTTATCTATATATGACGTATTGGAAGTACAAAAAGAAGAACTTG
pOPIN-His-SUMO-NIS1200-1500_rev	TAAACTGGTCTAGAAAGCTTACATATAAACCTTACTTATAAGAACTACATCAGGTAATTCA
pOPIN-His-SUMO-NIS1400-1834_fwd	GCGAACAGATCGGTGGTGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAA
pOPIN-His-SUMO-NIS1400-1834_rev	TAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-His-SUMO-NIS1500-1834_fwd	GCGAACAGATCGGTGGTGATGATAGTAAGCCTTCATTTGGATATTATAGTAATAATTTGAAAGAT
pOPIN-His-SUMO-NIS1500-1834_rev	TAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-His-SUMO-NIS800-1100_fwd	GCGAACAGATCGGTGGTTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-His-SUMO-NIS800-1100_rev	TAAACTGGTCTAGAAAGCTTAAATTCCTGCTAAAGGAACTAAAAGTATACTAAATCCA
pOPIN-His-SUMO-NIS800-1400_fwd	GCGAACAGATCGGTGGTTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-His-SUMO-NIS800-1400_rev	TAAACTGGTCTAGAAAGCTTAATTACTTCCATTTACCTCACCAGAAAAATTAATCTC
pOPIN-His-SUMO-NIS800-1834_fwd	GCGAACAGATCGGTGGTTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-His-SUMO-NIS800-1834_rev	TAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-His-SUMO-NIS900-1200_fwd	GCGAACAGATCGGTGGTTTTATTAATAAAGAAACTGGAGAATCTATATTTGTAGAAACTGAAAAAAC
pOPIN-His-SUMO-NIS900-1200_rev	TAAACTGGTCTAGAAAGCTTAGTGTGGCTCTCTATATGTTATTGATGGTGCT
pOPIN-NHis-3C-NIS1100-1400_fwd	GAAGTTCTGTTTCAGGGTCCCTCAGCAGGTATACCAAGCTTAGTAAACAATG

pOPIN-NHis-3C-NIS1100-1400_rev	CAAACTGGTCTAGAAAGCTTAATTACTTCCATTTACCTCACCAGAAAAATTAATCTC
pOPIN-NHis-3C-NIS1200-1500_fwd	GAAGTTCTGTTTCAGGGTCCCTTATCTATATATGACGTATTGGAAGTACAAAAAGAAGAACTTG
pOPIN-NHis-3C-NIS1200-1500_rev	CAAACTGGTCTAGAAAGCTTACATATAAACCTTACTTATAAGAACTACATCAGGTAATTCA
pOPIN-NHis-3C-NIS1400-1834_fwd	GAAGTTCTGTTTCAGGGTCCCGGATTTGTTTCTTTAACATTTTCAATTTTAGAAGGAATAAA
pOPIN-NHis-3C-NIS1400-1834_rev	CAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-NHis-3C-NIS1500-1834_fwd	GAAGTTCTGTTTCAGGGTCCCGATGATAGTAAGCCTTCATTTGGATATTATAGTAATAATTTGAAAGAT
pOPIN-NHis-3C-NIS1500-1834_rev	CAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-NHis-3C-NIS800-1100_fwd	GAAGTTCTGTTTCAGGGTCCCTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-NHis-3C-NIS800-1100_rev	CAAACTGGTCTAGAAAGCTTAAATTCCTGCTAAAGGAACTAAAAGTATACTAAATCCA
pOPIN-NHis-3C-NIS800-1400_fwd	GAAGTTCTGTTTCAGGGTCCCTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-NHis-3C-NIS800-1400_rev	CAAACTGGTCTAGAAAGCTTAATTACTTCCATTTACCTCACCAGAAAAATTAATCTC
pOPIN-NHis-3C-NIS800-1834_fwd	GAAGTTCTGTTTCAGGGTCCCTTATTACAAGAAATTAGAAATAATTCTAATTCAAGTGATATTGAACTAGAAGAA
pOPIN-NHis-3C-NIS800-1834_rev	CAAACTGGTCTAGAAAGCTTATCCAGATACCATCATTCCAAAGTTATTAATATAAAATTTCTCA
pOPIN-NHis-3C-NIS900-1200_fwd	GAAGTTCTGTTTCAGGGTCCCTTTATTAATAAAGAAACTGGAGAATCTATATTTGTAGAAACTGAAAAAAC
pOPIN-NHis-3C-NIS900-1200_rev	CAAACTGGTCTAGAAAGCTTAGTGTGGCTCTCTATATGTTATTGATGGTGCT

Table 36: Overview of oligonucleotides used for traditional cloning of TcdB constructs in this study.

Oligonucleotide	Sequence 5’ → 3’
540-2366NdeI_fw	GCCCGCCATATGCTTCTACTA
540-2366BamHI_rv	GTCCCGGGATCCCACTAATCA
800-1834NdeI_fw	GCCCGCCATATGAATAATGTT
800-1834BamHI_rv	GTCCTGGGATCCATCCAGATA
fl_TcdB_Xho_rv	GTCCCGCTCGAGATCCAGATA
800-1400_Nde_fw	CGCGCCCCGCATATGTTATTACAAGAA
800-1400_Xho_rv	CTGGTCCGCCTCGAGATTACTTCCATT
1401-1834_Nde_fw	GCCCCGCATATGGGATTTGTT
1401-1834_Xho_rv	GTCCTCCTCGAGATCCAGATA
Nde_fw1349-1811	GCCGCGCATATGGATAATGTTGTGAGAGAT
Xho_rv1349-1811	GCACCATAACTCGAGTTAATAGCCAATCAATCC
TcdB_C698>Ala	GAAATAAATTTATTAGGAGCGAATATGTTTAGCTACTCTATCAACGTA
TcdB_C698>Ser	GAAATAAATTTATTAGGAAGTAATATGTTTAGCTACTCTATCAACGTA

Table 37: Overview of oligonucleotides for SLIC cloning of synthetic TcdB constructs, prepared by Konrad Büssow (HZI, Brunswick, Germany).

Oligonucleotide	Sequence 5' → 3'
TcdB-TDCP-fwd	GAAGTTCTGTTTCAGGGTACCGAAGATGATAATCTGGATTTTAGCCAGAATATTG
TcdB-TDCP-RevM	TTAAACTGGTCTAGAAAGCTTCAACCGCTCACCATCATGCC
TcdB-TDCP-RevF	ACAAACTGGTCTAGAAAGCTTCAACCGCTCACCATCATGCC
TcdB-TD-Fwd	GAAGTTCTGTTTCAGGGTACCCTGCTGCAAGAAATTCGTAATAATAGC
TcdB-fl-Fwd	GAAGTTCTGTTTCAGGGTACCAGCCTGGTTAATCGTAAACAGC
TcdB-fl-RevM	TTAAACTGGTCTAGAAAGCTTCATTCACTAATCACCAGCTGTGCG
TcdB-fl-RevF	ACAAACTGGTCTAGAAAGCTTCATTCACTAATCACCAGCTGTGCG

Table 38: Overview of plasmids used and created in this study.

Plasmid	Source	Application
pHIS1522-TcdB-543-2366	Ralf Gerhard, MHH, Hannover	Protein production
pHIS1522-TcdB-543-2366(C698A)	Ralf Gerhard, MHH, Hannover	Protein production
pOPIN-M-TcdB-800-1100	this study (DPF)	Protein production
pOPIN-M-TcdB-800-1400	this study (DPF)	Protein production
pOPIN-M-TcdB-800-1834	this study (DPF)	Protein production
pOPIN-M-TcdB-900-1200	this study (DPF)	Protein production
pOPIN-M-TcdB-1000-1300	this study (DPF)	Protein production
pOPIN-M-TcdB-1100-1400	this study (DPF)	Protein production
pOPIN-M-TcdB-1200-1500	this study (DPF)	Protein production
pOPIN-M-TcdB-1300-1600	this study (DPF)	Protein production
pOPIN-M-TcdB-1400-1700	this study (DPF)	Protein production
pOPIN-M-TcdB-1400-1834	this study (DPF)	Protein production
pOPIN-M-TcdB-1500-1834	this study (DPF)	Protein production
pOPIN-S3C-TcdB-800-1100	this study (DPF)	Protein production
pOPIN-S3C-TcdB-800-1400	this study (DPF)	Protein production
pOPIN-S3C-TcdB-800-1834	this study (DPF)	Protein production
pOPIN-S3C-TcdB-900-1200	this study (DPF)	Protein production
pOPIN-S3C-TcdB-1000-1300	this study (DPF)	Protein production
pOPIN-S3C-TcdB-1100-1400	this study (DPF)	Protein production
pOPIN-S3C-TcdB-1200-1500	this study (DPF)	Protein production
pOPIN-S3C-TcdB-1300-1600	this study (DPF)	Protein production
pOPIN-S3C-TcdB-1400-1700	this study (DPF)	Protein production
pOPIN-S3C-TcdB-1400-1834	this study (DPF)	Protein production
pOPIN-S3C-TcdB-1500-1834	this study (DPF)	Protein production
pOPIN-F-TcdB-800-1100	this study (DPF)	Protein production
pOPIN-F-TcdB-800-1400	this study (DPF)	Protein production

pOPIN-F-TcdB-800-1834	this study (DPF)	Protein production
pOPIN-F-TcdB-900-1200	this study (DPF)	Protein production
pOPIN-F-TcdB-1000-1300	this study (DPF)	Protein production
pOPIN-F-TcdB-1100-1400	this study (DPF)	Protein production
pOPIN-F-TcdB-1200-1500	this study (DPF)	Protein production
pOPIN-F-TcdB-1300-1600	this study (DPF)	Protein production
pOPIN-F-TcdB-1400-1700	this study (DPF)	Protein production
pOPIN-F-TcdB-1400-1834	this study (DPF)	Protein production
pOPIN-F-TcdB-1500-1834	this study (DPF)	Protein production
pOPIN-F-TcdB-543-2366	this study (DPF)	Protein production
pOPIN-G-TcdB-1400-1834	this study (DPF)	Protein production
pOPIN-G-TcdB-800-1400	this study (DPF)	Protein production
pOPIN-G-TcdB-800-1834	this study (DPF)	Protein production
p10$-TcdB-1349-1811	this study	Protein production
pOPIN-M-TcdB1490-1848 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPIN-M-TcdB1-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU23-TcdB1490-1848 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU23-TcdB1-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPIN-J-TcdB1490-1848 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPIN-J-TcdB1-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPIN-M-TcdB-1-2366 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPIN-M-TcdB-543-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPIN-M-TcdB-800-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU10-TcdB-1-2366 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU10-TcdB-800-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU11-TcdB-1-2366 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU11-TcdB-543-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU11-TcdB-800-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU23-TcdB-1-2366 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU23-TcdB-543-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pPEU23-TcdB-800-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINJ-TcdB-1-2366 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINJ-TcdB-543-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINJ-TcdB-800-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINF-TcdB-1-2366 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINF-TcdB-543-1834	this study, Konrad Büssow (HZI)	Protein production

(synthetic TcdB)		
pOPINF-TcdB-800-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINS3C-TcdB-1-2366 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINS3C-TcdB-800-1834 (synthetic TcdB)	this study, Konrad Büssow (HZI)	Protein production
pOPINM-CD1363	this study (DPF)	Protein production
pOPINM-CD1364	this study (DPF)	Protein production
pOPINM-CD1365	this study (DPF)	Protein production
pOPINM-CD1366	this study (DPF)	Protein production
pOPINM-CD1367	this study (DPF)	Protein production
pOPINM-CD1368	this study (DPF)	Protein production
pOPINM-CD1369	this study (DPF)	Protein production
pOPINM-CD1370	this study (DPF)	Protein production
pOPINM-CD1371	this study (DPF)	Protein production
pOPINM-CD1372	this study (DPF)	Protein production
pOPINM-CD1373	this study (DPF)	Protein production
pOPINM-CD1374_85-536	this study (DPF)	Protein production
pOPINM-CD1374_911-1420	this study (DPF)	Protein production
pOPINM-CD1374fl	this study (DPF)	Protein production
p10$-CD10670	this study	Protein production
pMal-CD10670	this study	Protein production
pET-Sumo-CD10670	this study	Protein production
pET-Sumo-CD12190	this study	Protein production
p10$-CD18230	this study	Protein production
pET-Sumo-CD18230	this study	Protein production
p10$-CD25890	this study	Protein production
pET-Sumo-CD25890	this study	Protein production
pET-Sumo-CD25890-SER1	this study	Protein production
pMal-CD27520	this study	Protein production
pET-Sumo-CD27520	this study	Protein production

Table 39: Overview of produced constructs in *E. coli* with expression conditions used in this study.

Protein	*E. coli* strain	Expression conditions	Plasmid
His_6-Sumo-TcdB800-1834	BL21CodonPlus(DE3)RIL	4L LB medium, 0.5 mM IPTG, 20 °C, 16 h	pOPINS3C-TcdB800-1834
His_6-Sumo-TcdB800-1400	BL21CodonPlus(DE3)RIL	4L LB medium, 0.5 mM IPTG, 20 °C, 16 h	pOPINS3C-TcdB800-1400
His_6-Sumo-TcdB1500-1834	BL21CodonPlus(DE3)RIL	4L LB medium, 0.5 mM IPTG, 20 °C, 16 h	pOPINS3C-TcdB1500-1834
His_6-MBP-TcdB800-1100	BL21CodonPlus(DE3)RIL	4L autoinduction medium, 25 °C, 24 h	pOPINM-TcdB800-1100
His_6-MBP-TcdB1500-1834	BL21CodonPlus(DE3)RIL	4L autoinduction medium, 25 °C, 24 h	pOPINM-TcdB1500-1834

His_6-T7-Lysozyme-TcdB1349-1811	BL21CodonPlus(DE3)RIL	8L LB medium, 0.5 mM IPTG, 20 °C, 16 h	p10$-TcdB1349-1811
His_6-MBP-CD1363	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1363
His_6-MBP-CD1364	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1364
His_6-MBP-CD1367	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1367
His_6-MBP-CD1368	Rosetta2(DE3)	2L TB medium, 20 °C, 16 h, 0.5 mM IPTG	pOPINM-CD1368
His_6-MBP-CD1369	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1369
His_6-MBP-CD1370	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1370
His_6-MBP-CD1371	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1371
His_6-MBP-CD1372	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1372
His_6-MBP-CD1373	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1373
His_6-MBP-CD1374 (85-536)	BL21CodonPlus(DE3)RIL	2L autoinduction medium, 25 °C, 24 h	pOPINM-CD1374 (85-536)
His_6-MBP-CD1374 (911-1420)	Rosetta2(DE3)	2L TB medium, 20 °C, 16h, 1 mM IPTG	pOPINM-CD1374 (911-1420)
His_6-MBP-TcdB800-1834 (synthetic TcdB)	Rosetta2(DE3)	4L TB medium, 1 mM IPTG, 20 °C, 16 h	pOPINM-TcdB800-1834 (synthetic TcdB)
His_6-MBP-TcdB543-1834 (synthetic TcdB)	Rosetta2(DE3)	8L TB medium, 1 mM IPTG, 20 °C, 16 h	pOPINM-TcdB543-1834 (synthetic TcdB)
His_6-MBP-TcdB1490-1848 (synthetic TcdB)	BL21CodonPlus(DE3)RIL	8L TB medium, 1 mM IPTG, 20 °C, 16 h	pOPINM-TcdB1490-1848 (synthetic TcdB)
Strep-TcdB1490-1848 (synthetic TcdB)	BL21CodonPlus(DE3)RIL	8L TB medium, 1 mM IPTG, 20 °C, 16 h	pPEU23-TcdB1490-1848 (synthetic TcdB)
His_6-GST-TcdB1490-1848 (synthetic TcdB)	BL21CodonPlus(DE3)RIL	8L TB medium, 1 mM IPTG, 20 °C, 16 h	pOPINJ-TcdB1490-1848 (synthetic TcdB)
His_6-MBP-TcdB1-1834 (synthetic TcdB)	BL21CodonPlus(DE3)RIL	8L LB medium, 1 mM IPTG, 20 °C, 16 h	pOPINM-TcdB1-1834 (synthetic TcdB)
Strep-TcdB1-1834 (synthetic TcdB)	BL21CodonPlus(DE3)RIL	8L TB medium, 1 mM IPTG, 37 °C, 16 h	pPEU23-TcdB1-1834 (synthetic TcdB)
His_6-GST-TcdB1-1834 (synthetic TcdB)	BL21CodonPlus(DE3)RIL	8L TB medium, 1 mM IPTG, 20 °C, 16 h	pOPINJ-TcdB1-1834 (synthetic TcdB)
His_6-T7-Lysozyme-CD2589	BL21CodonPlus(DE3)RIL	4L TB medium, 1 mM IPTG, 20 °C, 16 h	p10$-CD2589
His_6-Sumo-CD1067	Rosetta2(DE3)	4L TB medium, 1 mM IPTG, 20 °C, 3 h	pET-Sumo-CD1067
His_6-Sumo-CD2752	BL21CodonPlus(DE3)RIL	4L TB medium, 1 mM IPTG, 20 °C, 16 h	pET-Sumo-CD2752
His_6-MBP-CD2752	Rosetta2(DE3)	4L TB medium, 1 mM IPTG, 20 °C, 16 h	pMal-CD2752
His_6-MBP-CD1067	Rosetta2(DE3)	4L TB medium, 1 mM IPTG, 20 °C, 16 h	pMal-CD1067
His_6-Sumo-CD1219	Rosetta2(DE3)	4L TB medium, 1 mM IPTG, 20 °C, 16 h	pET-Sumo-CD1219

His_6-Sumo-CD1823	BL21CodonPlus(DE3)RIL	4L TB medium, 1 mM IPTG, 20 °C, 16 h	pET-Sumo-CD1823
His_6-Sumo-CD2589 His_6-Sumo-CD2589-SER1	Rosetta2(DE3)	4L TB medium, 1 mM IPTG, 20 °C, 16 h	pET-Sumo-CD2589 pET-Sumo-CD2589-SER1
His_6-T7-Lysozyme-CD1823	BL21CodonPlus(DE3)RIL	4L TB medium, 1 mM IPTG, 20 °C, 16 h	p10$-CD1823
His_6-T7-Lysozyme-CD1067	Rosetta2(DE3)	4L TB medium, 1 mM IPTG, 20 °C, 16 h	p10$-CD1067
His_6-T7-Lysozyme-CD2752	BL21CodonPlus(DE3)RIL	8L TB medium, 1 mM IPTG, 20 °C, 16 h	p10$-CD2752

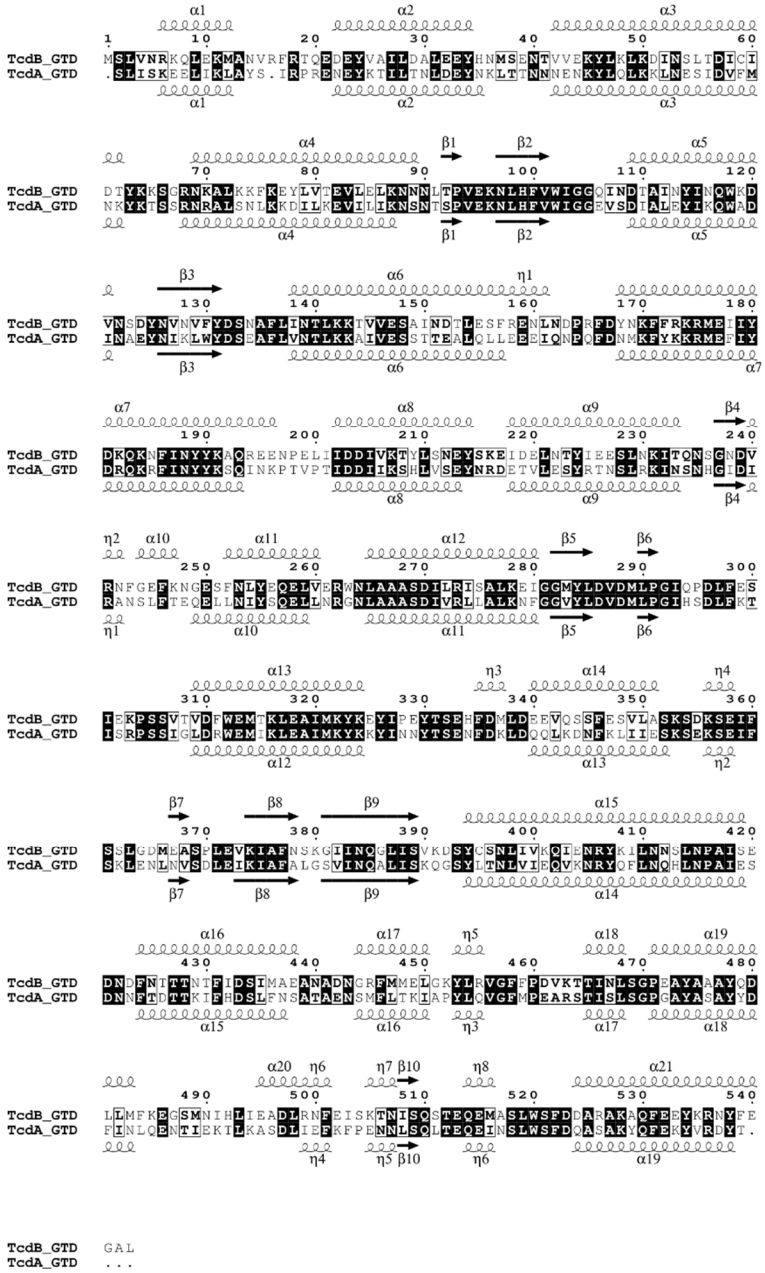

Figure 81: Structure-based sequence alignment of the glucosyltransferase domain of TcdA (PDB-Code 3SRZ) and TcdB (PDB-Code 2BVL). Secondary structure elements of TcdA-GTD are depicted under the alignment, TcdB-GTD structure is illustrated above the alignment. Conserved residues are highlighted by a black box. α-helices are depicted by spirals, β-strands by arrows. The secondary structure elements are numbered consecutively.

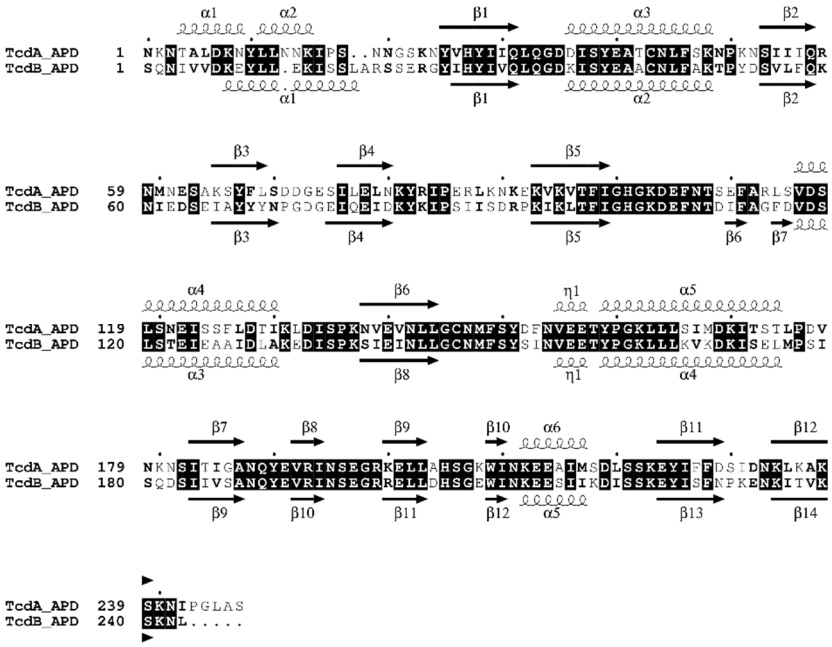

Figure 82: Structure-based sequence alignment of the autoprotease domain of TcdA (PDB-Code 3HO6) and TcdB (PDB-Code 3PEE). Secondary structure elements of TcdA-APD are depicted above the alignment, TcdB-APD structure is illustrated under the alignment. Conserved residues are highlighted by a black box. α-helices are depicted by spirals, β-strands by arrows. The secondary structure elements are numbered consecutively.

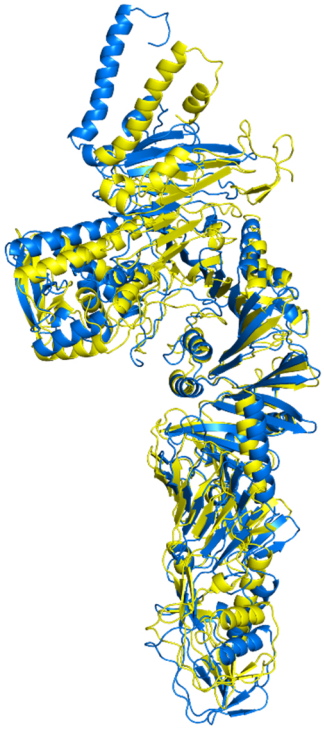

Figure 83: Structural alignment of the intermediate translocation domain of TcdA (PDB-Code 4R04, in blue) and the Phyre model of the TcdB translocation domain (yellow).

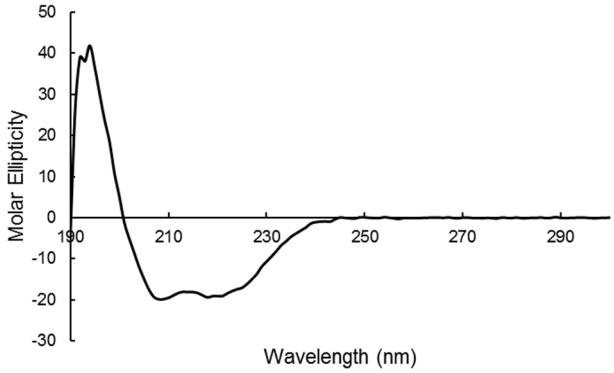

Figure 84: CD spectrum of native CD2589 in 20 mM Tris/HCl pH 8.0, 150 mM NaCl

12. Danksagung

Ich möchte hiermit allen Leuten danken, die mich während der Zeit meiner Doktorarbeit unterstützt haben und zur erfolgreichen Abgabe dieser Arbeit beigetragen haben:

Danke, Wulf! Dafür, dass ich meine Doktorarbeit in deiner Arbeitsgruppe schreiben durfte, dass ich parallel zum Toxin-Projekt an weiteren Proteinen arbeiten durfte, aber v.a. für die Unterstützung aber auch den nötigen Freiraum während meiner gesamten Arbeit.

Danke an Prof. Dr. Michael Hust für die Übernahme des Vorsitzes meiner Prüfungskommission und an Prof. Dr. Michael Steinert für die Übernahme der Zweitkorrektur.

Danke auch an die Mitglieder meines Thesis Committees Prof. Dr. Ralf Gerhard und Dr. Andrea Scrima für die angeregten Diskussionen, unterhaltsamen Meetings und die Hilfe bei sämtlichen Fragestellungen. Danke auch besonders an Ralf für das Strep-TcdB Konstrukt, das dann letztendlich den Durchbruch bei der TcdB-Expression gegeben hat.

Danke an alle Mitarbeiter der Arbeitsgruppen SFPR, Ex-SBAU und RPEX, die das Arbeiten im Labor erleichtert und mit hilfreichen Ratschlägen versehen haben. Danke PP1 (Postdoc-Peer), PP2 (Postdoc-Poppe), PP3 (Postdoc-Pippel), Stefan, Kwang-Hoon, Yafei, Thomas, Ute, Claudia, Angelika, Simone für die spitzen Arbeitsatmosphäre!

Danke an Mihai, der mich v.a. zu Beginn meiner Doktorarbeit unterstützt hat und mir mein Hobby-/Wochenend-Projekt der Diffocine überlassen hat. Besonderer Dank gilt auch Jan, der mir bei der Strukturlösung des Diffocin-Tube-Proteins geholfen und den großen Teil des Paper Schreibens übernommen hat.

Danke an Christina, meine Betreuerin meiner Masterarbeit, die auch während meiner Doktorarbeit immer ein offenes Ohr für Fragen hatte und mir immer geholfen hat. Ich habe so viel von dir gelernt, danke! Danke aber auch den anderen Senior-PhDs Allegra, Bockto (alias Bobi) und Christian, die den Anfang hier in Braunschweig erheblich erleichtert haben und immer geholfen haben, wo es nur ging.

Danke an alle Arbeitskollegen, die zu richtig guten Freunden wurden. Maurice, danke für die vielen unterhaltsamen Kaffeepausen, Ausflüge in den Zoo, Besuche auf der Reeperbahn, durchgefeierte Nächte in Braunschweig usw. Danke Björn, für diverse Wrestling-Späße, Stadion-Besuche, Party-Abende, einfach für alles! Danke auch an meine Bürokollegen Juliane, Flo & Kevin, die den Büro-Alltag sehr unterhaltsam gestaltet haben.

Danke an die Franken-Connection Flo und Moni alias Floni. Ohne euch wäre das alles nichts geworden. Danke für den täglichen Wahnsinn auf Arbeit, aber auch außerhalb der Arbeit und den Spaß, den wir zusammen hatten. Es war spitze!

Danke auch an meine Familie Linda, Mutti, Oma & Opa und Dominik. Danke, dass ihr mich immer unterstützt habt, mit aufmunternden und lustigen Nachrichten, spontanen Besuchen und Telefonaten in schlaflosen Nächten. Zuhause ist da, wo man sich vermisst und Mutti ist!